Stochastic basins of attraction and generalized committor functions

Michael Lindner
Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany and Department of Mathematics, Humboldt University, Rudower Chaussee 25, 12489 Berlin, Germany

Frank Hellmann
Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany

(Received 6 March 2019; published 19 August 2019)

We study two generalizations of the basin of attraction of a stable state, to the case of stochastic dynamics, arbitrary regions, and finite-time horizons. This is done by introducing generalized committor functions and studying soujourn times. We show that the volume of the generalized basin, the basin stability, can be efficiently estimated using Monte Carlo–like techniques, making this concept amenable to the study of high-dimension stochastic systems. Finally, we illustrate in a set of examples that stochastic basins efficiently capture the realm of attraction of metastable sets, which parts of phase space go into long transients in deterministic systems, that they allow us to deal with numerical noise, and can detect the collapse of metastability in high-dimensional systems. We discuss two far-reaching generalizations of the basin of attraction of an attractor. The basin of attraction of an attractor are those states that eventually will get to the attractor. In a generic stochastic system, all regions will be left again; no attraction is permanent. To obtain the equivalent of the basin of attraction of a region we need to generalize the notion to cover finite-time horizons and finite regions. We do so by considering soujourn times, the fraction of time that a trajectory spends in a set, and by generalizing committor functions which arise in the study of hitting probabilities. In a simplified setting we show that these two notions reduce to the normal notions of the basin of attraction in the appropriate limits. We also show that the volume of these stochastic basins can be efficiently estimated for high-dimensional systems at computational cost comparable to that for deterministic systems. To fully illustrate the properties captured by the stochastic basins, we show a set of examples ranging from simple conceptual models to high-dimensional inhomogeneous oscillator chains. These show that stochastic basins efficiently capture metastable attraction, the presence of long transients, that they allow us to deal with numerical and approximation noise, and can detect the collapse of metastability with increasing noise in high-dimensional systems.

DOI: 10.1103/PhysRevE.100.022124

I. INTRODUCTION

A key question in the study of dynamical systems is the question of their long-time behavior. For deterministic systems this is captured in the notion of the attractors of the system and their basin of attraction. Informally, an attractor in a dynamical system is a minimal forward-invariant set A which admits a larger set V, such that all states in V converge to A. Its basin of attraction B_A is the set of all such convergent states. An important quantity when concerned with multistable systems is the probability that a trajectory converges to a specific attractor after an initial, possibly large perturbation. Thus, the basin stability [1,2] of A is defined as the volume of its basin of attraction under a given perturbation measure. An attractor A with nonzero basin stability is called Milnor attractor [3]. So far basin stability has been studied almost exclusively in the context of deterministic dynamics, since if the system is stochastic, then the topological notions of attractor and basin of attraction are no longer immediately meaningful.

The study of the long-term behavior of stochastic systems instead focuses on the question of its invariant measure, which replace the role of invariant sets. Ergodic theory [4,5] asks under which conditions and in which sense initial conditions converge to an invariant measure, if this measure is unique, and what its characteristic properties are. In other words, ergodic theory is concerned with the asymptotic behavior of measure-preserving dynamical systems. However, in many stochastic systems of interest, there is only one invariant measure. For such systems phenomena of multistability are not visible in the infinite-time behavior limit.

On finite timescales metastable states [6] and almost-invariant sets [7–10] correspond in many ways to attractors and invariant sets, that is they generalize an attractor's transient properties, the influence it exerts on trajectories in its neighborhood. When a system is a weak perturbation of a multistable deterministic system, the basins of attraction get perturbed into metastable sets in backward time. One can also consider sets which are metastable in both time directions, or the more general notion of sets that stay coherent under nonautnomous dynamics [11,12].

This paper explores two quantities that directly generalize the notion of the basin of attraction of an attractor, by defining the basin of attraction of a generic set at finite-time horizon. Rather than starting from invariant sets we will consider the notion of attraction to specified regions. Importantly, these

generalized basins remain amenable to Monte Carlo study in high-dimensional systems. Just as with basin stability for deterministic systems, there exists a probabilistic formulation of the volume of these basins that enable the formulation of straightforward efficient estimators. In particular this means that the basin stability of an attractive set can be evaluated without explicitly determining the set itself. The cost of evaluating it to a fixed accuracy scales with the cost of sampling a trajectory of the system. We will see this in an application to a 32-dimensional SDE.

Concretely, our generalized stochastic basins are based on mean sojourn times [13] and on hitting probabilities [14], which are also known as committor functions [15], and are immediately applicable to stochastic systems. Committor functions have previously been used as a tool to study deterministic basins and to optimize basin stability in [16–19].

For the case of finite state spaces and discrete time we will show that both quantities we explore recover the usual notions of the basin of attraction in the limit of the time horizon going to infinity. Thus, they serve as genuine finite-time horizon generalization of the basin of attraction in this case. We expect this to hold in rather more general contexts as well. We show in a set of examples that they capture the expected properties of metstable sets but also that they can be used to study transient phenomena not typically captured otherwise. One important observation is that the notion of attraction now contains two aspects, bringing states to the set and preventing them from leaving. Finally, we demonstrate an example of treating a high-dimensional systems by sampling the trajectory space in an efficient manner.

Recently, Serdukova *et al.* [20] also proposed the notion of stochastic basins of attraction for deterministic systems with noise. Their definition is based on studying the escape probability from the basin of attraction of the underlying deterministic system. In contrast, our definitions work without assuming an underlying deterministic system or knowledge of its basin structure. This allows the construction of efficient estimators for the associated basin stability in high-dimensional systems. We will further discuss the relationship to this work in the conclusion.

II. MARKOV CHAINS

We will introduce the stochastic basins of attraction using stochastic, discrete time, discrete space dynamical systems, that is finite-dimensional Markov chains X_k. While most of the discussion will focus on this setting (mostly for mathematical convenience), the probabilistic nature of our fundamental definitions makes it straightforward to transport the notions to other systems. This will be done in Sec. V where we will discuss how properties of the stochastic basins defined here can be studied in high-dimensional continuous time, continuous space systems. Further, a Markov chain can approximate a smooth deterministic system, too, for example, through Ulam's method. This is most easily seen in the Transfer operator approach to dynamical systems. For the reader unfamiliar with this approach, Appendix A contains a detailed discussion.

We will briefly review two notions on Markov chains, mean sojourn times and committor functions, that are fundamental to our approach.

A *Markov chain* [14] $(X_k)_{k\in\mathbb{N}}$ is a stochastic process on a discrete state space $\mathcal{X}$, such that X_k is a random variable with values in $\mathcal{X}$ for all $k \in \mathbb{N}$ and such that the Markov property is satisfied, i.e., for all $k \in \mathbb{N}$:

$$\begin{aligned}\mathbb{P}[X_{k+1} = i \mid X_k = i_k, \ldots, X_0 = i_0] \\ = \mathbb{P}[X_{k+1} = i \mid X_k = i_k],\end{aligned} \tag{1}$$

where $i, i_0, \ldots, i_k$ are arbitrary elements of $\mathcal{X}$. That is, the probability for the next state only depends on the current state, rather than on the history. A Markov chain is called *homogeneous* or *stationary* if

$$\mathbb{P}[X_{k+1} = i \mid X_k = j] = \mathbb{P}[X_1 = i \mid X_0 = j] \quad \forall k \in \mathbb{N}. \tag{2}$$

There is a one-to-one correspondence between homogeneous Markov chains on a finite-state space $\mathcal{X} = \{1, \ldots, n\}$ and stochastic matrices $M \in \mathbb{R}^{n\times n}$ by setting

$$\mathbb{P}[X_1 = j \mid X_0 = i] = M_{ij}. \tag{3}$$

A *probability distribution vector* $\rho \in \mathbb{R}^n$ is a nonnegative vector, such that $\sum_{i=1}^n \rho_i = 1$. If $\rho := \rho(0)$ specifies the initial distribution of the Markov chain, i.e., $\mathbb{P}[X_0 = i] = \rho_i$, then the kth step distribution $\rho(k)_i = \mathbb{P}[X_k = i \mid X_0 \sim \rho]$ can be computed as

$$\rho(k) = \rho^T M^k. \tag{4}$$

From now on we assume that $(X_k)_{k\in\mathbb{N}}$ is homogeneous.

A. Sojourn times

Let $A \subseteq \{1, \ldots, n\}$ be a subset of the state space. The *mean sojourn time* in A is the relative amount of time that the process spends in A. Let $\mathbb{1}_A(x) = 1$ if $x \in A$, else $\mathbb{1}_A(x) = 0$, denote the indicator function on A, then the mean sojourn time $\tau_s(A)$ along a trajectory is the random variable

$$\tau_s(A) := \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{1}_A(X_k). \tag{5}$$

The *expected mean sojourn time* or *EMS time* in A is

$$s(A) := \mathbb{E}[\tau_s(A)] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \rho^T M^k \mathbb{1}_A, \tag{6}$$

where ρ is the initial distribution of the Markov chain. Equation (6) holds since

$$\mathbb{E}[\mathbb{1}_A(X_k)] = \mathbb{P}[X_k \in A] = \sum_{i\in A} \rho(k)_i = \rho^T M^k \mathbb{1}_A. \tag{7}$$

The mean soujourn time is an infinite-time limit that only depends on the asymptotic distribution of the Markov chain. To study the transient behavior, we will consider the finite-horizon EMS time, that is, the fraction of a finite-time horizon N that the system spends in A:

$$s_N(A) := \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E}[\mathbb{1}_A(X_k)] = \frac{1}{N} \sum_{k=0}^{N-1} \rho^T M^k \mathbb{1}_A. \tag{8}$$

The finite-horizon EMS time is the expectation value of the operator

$$S_N(M) = \frac{1}{N}\sum_{k=0}^{N-1} M^k \ . \tag{9}$$

B. Committor functions

Committor functions give the probability that a system hits a region $A \subseteq \mathcal{X}$. This *absorption probability* q can be obtained as the minimal nonnegative solution of the system of equations [14]

$$\begin{aligned} Mq &= q \quad \text{on } \mathcal{X} \setminus A, \\ q &= 1 \quad \text{on } A. \end{aligned} \tag{10}$$

Similarly, for two disjoint sets A and B, the probability of not entering a set B before having visited A is given by the minimal nonnegative solution of

$$\begin{aligned} Mq &= q \quad \text{on } \mathcal{X} \setminus (A \cup B), \\ q &= 1 \quad \text{on } A, \\ q &= 0 \quad \text{on } B. \end{aligned} \tag{11}$$

The solution q is called *committor function*. An equivalent way of looking at the problem is to modify the process by adding two exit states Z_A, Z_B to the state space, such that the transition probability from any state in A to the exit state Z_A is 1 and equally for B and Z_B.

III. NEW CONCEPTS FOR COMMITTOR FUNCTIONS

In some sense the committor of a set A is the most straightforward generalization of the notion of a basin of attraction. However, for generic stochastic systems, and generic regions A, the probability to eventually enter A is 1. This is why the relative committor of hitting A before B is usually studied. To capture the ability of a region A to attract and retain trajectories we thus introduce generalized notions of committors.

In this section we introduce two new generalizations of committor functions, fuzzy committors and ε-committors, and study some of their basic properties. In the next section we will see that the latter, as well as finite-horizon EMS times, are natural generalizations of the notion of the basin of attraction.

A. Fuzzy committors

When we introduce exit states we are free to choose arbitrary transition probabilities to the exit states. Assume we have two transition probability distributions p^1, p^2 into exit states Z_1, Z_2. Additionally, we require that $0 \leqslant p_i^1 + p_i^2 \leqslant 1$ and $p_i^1, p_i^2 \geqslant 0$ for all $i \in \mathcal{X}$. To obtain the probability of being absorbed into Z_1 we introduce the transition matrix of the augmented process $\hat{X}$,

$$\hat{Q} = \left(\begin{array}{c|cc} \hat{M} & p^1 & p^2 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right), \tag{12}$$

where $\hat{M}_{ij} := M_{ij} \cdot (1 - p_i^1 - p_i^2)$ and obtain the minimal, nonnegative solution $\hat{q} = (q, 1, 0)$ to the system

$$\begin{aligned} \hat{Q}\hat{q} &= \hat{q} \quad \text{on } \mathcal{X}, \\ \hat{q} &= 1 \quad \text{on } Z_1, \\ \hat{q} &= 0 \quad \text{on } Z_2. \end{aligned} \tag{13}$$

This can be rephrased as

$$\begin{aligned} &\hat{M}q + p^1 = q, \\ \Leftrightarrow\ &(I - \hat{M})q = p^1, \end{aligned} \tag{14}$$

where we denote the absorption probability into exit state Z_1 by q. Note, however, that $\hat{M} := \hat{M}(p^1, p^2) := \hat{M} \cdot \mathrm{diag}(\mathbb{1} - p^1 - p^2)$ depends on p^1, p^2. If we set $p^1 = \mathbb{1}_A$, $p^2 = 0$, then we obtain Eq. (10), and if we set $p^1 = \mathbb{1}_A$, $p^2 = \mathbb{1}_B$, then we obtain Eq. (11).

The probability distribution q will be referred to as *fuzzy committor with respect to* p^1 *and* p^2. The term *fuzzy* refers to the notion of fuzzy sets described by affiliation functions like p^1 and p^2, generalizing the idea of "crisp" sets commonly described by binary indicator functions.

B. ε-committors

We now come to the key concept introduced in this paper to address stability questions at finite timescales.

Consider a Markov chain that has probability ε of being absorbed into a unique exit state at every time step and uniformly on its state space. Define a random variable T_ε as the time step when a trajectory is absorbed into the exit state. Thus, $T_\varepsilon - 1$ is the time the system spends in the original state space. Clearly, $\mathbb{P}[T_\varepsilon = 0] = 0$ and the probability that a trajectory hits the exit state at time step $k > 0$ is

$$\mathbb{P}[T_\varepsilon = k] = \varepsilon(1-\varepsilon)^{k-1} \quad \forall k \geqslant 1. \tag{15}$$

The expected value of T_ε is

$$\mathbb{E}[T_\varepsilon] = \sum_{k=0}^{\infty} k\, \mathbb{P}[T_\varepsilon = k] = \varepsilon \sum_{k=1}^{\infty} k(1-\varepsilon)^{k-1} = \cdots = \frac{1}{\varepsilon}, \tag{16}$$

and hence the inverse of the exit probability ε is the expected time horizon probed by such a modified process.

We want to know the probability that a given trajectory starting in state i moves toward a set A and stays there with respect to a finite timescale. Choose an exit probability $\varepsilon \in (0, 1]$ as the inverse of the timescale of interest and define $p^1 = \varepsilon\mathbb{1}_A$, $p^2 = \varepsilon(\mathbb{1} - \mathbb{1}_A)$ as exit probabilities for the fuzzy committors. Then Eq. (14) becomes

$$(I - (1-\varepsilon)M)q_\varepsilon(A) = \varepsilon\mathbb{1}_A. \tag{17}$$

The solution $q_\varepsilon(A) =: q_\varepsilon$ will be referred to as *ε-committor* of A, where we drop the argument A if it is clear from context. Existence and uniqueness of q_ε for $\varepsilon > 0$ follow by applying the Neumann inversion formula

$$q_\varepsilon(A) = \varepsilon\left[\sum_{k=0}^{\infty}(1-\varepsilon)^k M^k\right]\mathbb{1}_A. \tag{18}$$

The ε-committor again is defined in terms of an operator

$$C_\varepsilon(M) = \varepsilon \sum_{k=0}^{\infty} (1-\varepsilon)^k M^k. \tag{19}$$

Interpreted in terms of exit states, the ε-committor gives the probability to exit the system through region A. Intuitively it is clear that it will be large if the system spends a lot of time in A, that is, if it moves there quickly and stays there long relative to our chosen time horizon.

Note that we can replace the target set A characterized by an indicator function $\mathbb{1}_A$ with a generalized state described by any vector $v \in \mathbb{R}^n$, such that $\max_i |v_i| = 1$. The corresponding ε-committor is denoted as $q_\varepsilon(v)$. We will not pursue this possibility further in this paper though.

IV. ε-COMMITTORS AND EMS TIMES AS GENERALIZED BASINS

Both the finite-horizon EMS time and the ε-committors are generalizations of the basin of attraction of a system. To demonstrate this we make use of the following ergodic theorem:

Theorem IV.1. Let $M \in \mathbb{R}^{n\times n}$ be a stochastic matrix and let $Q \in \mathbb{R}^{n\times n}$ be invertible, such that $J = Q^{-1}MQ$ is the Jordan normal form of M, then

$$\lim_{N\to\infty} S_N[M] = \lim_{\varepsilon\to 0} C_\varepsilon[M] = P_{\mathrm{fix}(M)}, \tag{20}$$

where $P_{\mathrm{fix}(M)}$ is the projection onto $\mathrm{fix}(M) = \{v \in \mathbb{R}^n \mid v^T M = v^T\}$ given by

$$P_{\mathrm{fix}(M)} = Q^{-1} P_{\mathrm{fix}(J)} Q, \tag{21}$$

and $P_{\mathrm{fix}(J)}$ is an orthogonal projection.

The proof is given in Appendix B. The right eigenvectors corresponding to the v^T exist and are given by the corresponding columns of Q^{-1}. Thus, $P_{\mathrm{fix}(M)} = \sum_i w_i v_i^T$.

If M describes an underlying deterministic system, then the attractors A are exactly sets that satisfy $\mathbb{1}_A^T M = \mathbb{1}_A^T$, and these form a basis of the space of eigenvectors with eigenvalue 1. The corresponding right eigenvectors $\mathbb{1}_{B_A}$ are the basin structure (see Appendix A).

Thus, we can write the projector as a sum over attractors:

$$S_\infty[M] = C_0[M] = P_{\mathrm{fix}(M)} = \sum_A \mathbb{1}_{B_A} \mathbb{1}_A^T, \tag{22}$$

that is, a density that is in a particular basin get's projected onto the attractor. The asymptotic committor $q_0(A')$ for an attractor A' is simply given by $\mathbb{1}_{B_{A'}}$, and the EMS $s_\infty(A')$ for an initial distribution ρ is given by

$$s_\infty(A') = \rho \cdot q_0(A') = \rho \cdot \mathbb{1}_{B_{A'}}. \tag{23}$$

This demonstrates that both the finite horizon EMS and the ε-committor are genuine generalizations of the notion of the basin of attraction for this type of systems. For systems that can be approximated by Markov chains, for example, by means of Ulam's method, we expect similar results to hold.

The existence of a simple linear Eq. (17) for q_ϵ means the finite-time ϵ committor can be evaluated efficiently; however, the dimensionality of the Markov chain grows exponentially in the number of dimensions of the system approximated. In Sec. V we will therefore look at evaluating quantities like $\rho q_0(A')$ directly through its probability interpretation, rather than by approximating it with a finite-dimensional Markov chain. This Monte Carlo–based approach generalizes to high-dimensional systems.

We now turn to proving some more properties of the ε-committors themselves that elcuidate what exact properties of the system they capture.

The ε-committor is closely related to the expected time a trajectory spends in a given set A. To see this, define a random variable $\tau_\varepsilon = \tau_\varepsilon(A)$ as the total time the augmented process $\hat{X}$ including exit states spends in the set A, that is

$$\tau_\varepsilon = \sum_{k=0}^{\infty} \mathbb{1}_A(\hat{X}_k). \tag{24}$$

Then the expected value of τ_ε is

$$\begin{aligned}\mathbb{E}[\tau_\varepsilon] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \mathbb{1}_A(\hat{X}_k)\right] = \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{1}_A(\hat{X}_k)] = \sum_{k=0}^{\infty} \mathbb{P}[\hat{X}_k \in A] \\ &= \sum_{k=0}^{\infty} \mathbb{P}[X_k \in A](1-\varepsilon)^k. \end{aligned} \tag{25}$$

If we condition on an initial state $Y_0 = X_0 = i$, then Eq. (25) becomes

$$\begin{aligned}\mathbb{E}[\tau_\varepsilon \mid X_0 = i] &= \sum_{k=0}^{\infty} \mathbb{P}_i[X_k \in A] \cdot (1-\varepsilon)^k \\ &= \sum_{k=0}^{\infty} e_i^T M^k \mathbb{1}_A (1-\varepsilon)^k = \frac{1}{\varepsilon} q_{\varepsilon,i}, \end{aligned} \tag{26}$$

which relates the ε-committor to the expected time that the process spends in A before absorption. This could be small because the process does not reach A, and thus A is not attractive on this timescale, or because it quickly leaves A again, that is, A is not stable.

If the limit $\varepsilon \to 0$ of $\frac{1}{\varepsilon} q_{\varepsilon,i}$ exists, then it equals the expected time the original process X_k spends in A. The probability distribution function of $\tau_\varepsilon(A)$ for general absorbing Markov chains is derived in Csenki [21], Corollary 2.8.

Given the similarity in interpretation of the mean soujourn time and the ε-committor, as well as the fact that they have the same asymptotic limit, it is natural to ask if they generally give the same results. In Appendix C we obtain difference estimates between both quantities for specific metastable states. The difference vanishes for both small and for large times, and thus also for a sufficiently pronounced spectral gap. We will further explore the properties of the EMS and ε-committors with respect to metastable states in the examples.

V. GENERALIZED BASIN STABILITY

Crucially the mean soujourn time and the ε-committor are generalizations that allow for a notion of basin stability that can be evaluated efficiently by sampling. In the preceding sections we defined the concepts for discrete systems, and it is technically involved to generalize the discussion rigorously to continuous time and state space. In contrast, the sampling procedure we consider here immediately generalizes

to continuous time and state space. It also points toward a wider variety of "ε-committor-like functions" that might be of interest for further study.

Basin stability is the probability that a deterministic system returns to a desirable attractor after a perturbation. Typically the perturbations are described by a probability density on phase space $\rho_{\text{pert}}(x)$. The basin stability b is then simply given by the integral of the characteristic function of the basin B with respect to $\rho_{\text{pert}}(x)$:

$$b = \int \mathbb{1}_B(x)\rho_{\text{pert}}(x)dx. \tag{27}$$

This integral can be evaluated using Monte Carlo integration. Alternatively, we can interpret the sampling directly as a Bernoulli experiment, drawing initial conditions and observing whether or not the system returns to the attractor. Crucially, we do not need to know the shape of the basin to estimate b. The relative accuracy of the unbiased estimator $\hat{b}(N_b)$ obtained by sampling N_b trajectories is asymptotically small and independent of system details. Specifically the standard error of the estimator is given by

$$\sigma_{\hat{b}(N_b)} = \sqrt{\frac{\hat{b}(N_b)(1-\hat{b}(N_b))}{N_b}} + O\big(N_b^{-1}\big). \tag{28}$$

In the case of general, not necessarily deterministic dynamics, the generalized basin stability of a set A can be defined as

$$b_{\text{gen}} = \int q_{\text{gen}}(x)\rho_{\text{pert}}(x)dx, \tag{29}$$

where q_{gen} is the generalized membership function of the basin of A. In particular, we can choose q_{gen} to be the epsilon committor $q_\varepsilon(x)$ or the expected mean soujourn time $s_T(x)$.

A Monte Carlo estimation of this integral would be more expensive, as, for a stochastic system, $q_\varepsilon(x)$ or $s_T(x)$ can not be evaluated using only a single experiment. However, we can again design a Bernoulli experiment with expected probability b_{gen}. The experiment is as follows: Draw an initial condition from ρ_{pert}, run the system for a randomly chosen time t, and then check if it is in A at that time.

To see this, note that the generalized membership functions themselves have the interpretation as the probability of a Bernoulli experiment. They correspond to the probability to run to A in time t when starting from some initial condition x if we draw t from an appropriate choice distribution ρ_{run}. For $q_\varepsilon(x)$ we take the exit time distribution $\epsilon e^{-\epsilon t}$ as the run-time distribution ρ_{run}. This amounts simply to reinterpreting the exit from the system as run duration. For $s_T(x)$ we take the distribution of run times to be the equidistribution on the time interval $[0, T]$, so that the expectation value is equivalent to averaging in the time interval.

These definitions in terms of probabilities naturally extend to continuous times:

$$\begin{aligned} q_\varepsilon(x) &= \int dt\; p(x(t) \in A | x(0) = x)\, p(t_{\text{exit}} = t) \\ &= \int dt\; p(x(t) \in A | x(0) = x)\, \varepsilon e^{-\varepsilon t}, \\ s_T(x) &= \int dt\; p(x(t) \in A | x(0) = x)\, \frac{1}{T}\mathbb{1}_{[0,T]}(t). \end{aligned} \tag{30}$$

Now the integral in Eq. (29) is simply given by drawing the initial condition from ρ_{pert}:

$$\begin{aligned} b_{\text{gen}} &= \int dt \int dx\; p(x(t) \in A | x(0) = x)\, \rho_{\text{run}}(t)\, \rho_{\text{pert}}(x) \\ &= p(x(T) \in A | p(x(0)) = \rho_{\text{pert}}, p(T) = \rho_{\text{run}}). \end{aligned} \tag{31}$$

This is the probability of a Bernoulli experiment and thus can be studied by sampling again. We see immediately that this is true for a large class of such measures, namely, for all distributions of the evaluation time that are efficient to sample. Among these the two concepts developed in this paper are distinguished by taking the run time to be either given by a constant stopping rate or by a constant function.

This experiment will have the same variance of the estimator as the deterministic basin stability Eq. (28). Note further that after obtaining a sample of trajectories it is possible to evaluate the generalized basin stabilities for different sets on this sample, with each individual error given by Eq. (28). However, the errors will be correlated, making it hard to do statistics on the various measures thus obtained.

VI. EXAMPLES

We will illustrate the properties of the ε-committors using examples ranging from a low-dimensional Markov chain, through a high-dimensional Markov chain approximating a low-dimensional dynamical system, all the way to a high-dimensional stochastic differential equation.

A. Conceptual box models

The following simple systems are paradigmatic for the cases of metastability and long transients. They show that the notions introduced above accurately capture the basin of states that get attracted by metastable sets or into long transients.

Figure 1 shows the state-transition diagrams of two very simple Markov chains. One of them consists of two almost-invariant states M_1 and M_2 and provides a conceptual model of metastability in ergodic systems. The other one contains an attractor A and two transient states M_1 and M_2, where trajectories spend a long time before converging to A. The second model illustrates the concept of long transient states in dissipative systems. The transition matrices are

$$\begin{pmatrix} 1-\delta^2 & \delta^2 \\ \delta & 1-\delta \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1-\delta^2 & 0 & \delta^2 \\ 0 & 1-\delta & \delta \\ 0 & 0 & 1 \end{pmatrix}. \tag{32}$$

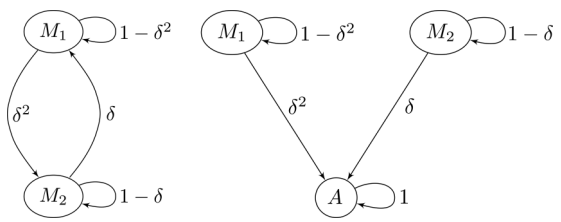

FIG. 1. Conceptual models of metastability and long transients.

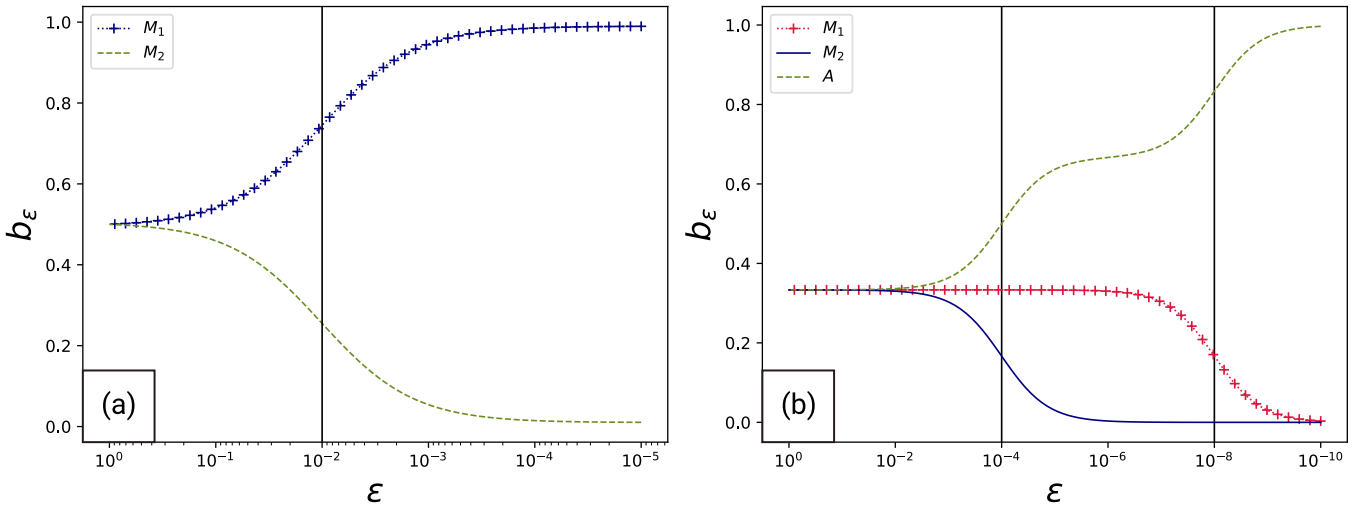

FIG. 2. ε-absorption stability for the model with almost-invariant states (a) and long transient states (b). The initially identical values converge asymptotically to the invariant distribution with decreasing ε. For the metastability model the invariant distribution on M_1 is close to 1 but strictly smaller. In the long transient model $b_\varepsilon(M_1)$ stays almost constant over a large interval of ε, since the leak rate δ^2 is very small. The horizontal lines indicate the value of δ, respectively δ^2.

We compute ε-absorption stability for these systems according to Eq. (17) by solving

$$b_\varepsilon(i) = \mathbb{1}^T \frac{\varepsilon}{n}(I - (1-\varepsilon)M)^{-1} e_i, \tag{33}$$

where M is the transition matrix and e_i is the standard basis vector corresponding to state $i = M_1, M_2, A$ and $n = 2, 3$. The parameter δ controls the timescales and is chosen to be 0.01 in the metastability model and 0.0001 in the long transients model. Figure 2 shows ε-absorption stability of the different states for varied ε.

The limits of $\varepsilon[I - (1-\varepsilon)M]^{-1}$ for ε to 0 are

$$\frac{1}{\delta^2+\delta}\begin{pmatrix}\delta & \delta^2\\ \delta & \delta^2\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}0 & 0 & 1\\ 0 & 0 & 1\\ 0 & 0 & 1\end{pmatrix}, \tag{34}$$

and it follows that the 0-absorption stability b_0 is

$$b_0 = \lim_{\varepsilon\to 0} b_\varepsilon = \frac{1}{\delta^2+\delta}\begin{pmatrix}\delta\\ \delta^2\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}0\\ 0\\ 1\end{pmatrix}, \tag{35}$$

which are just the invariant distributions. This shows that for the long transient model we recover the usual basin stability value in the ε to 0 limit. For the metastability model basin stability is not well-defined since trajectories never converge to an attractor. In this case ε-absorption stability converges to the invariant distribution.

Figure 2(b) shows that for finite ε the attraction of a region on that timescale is accurately captured. With ε between δ and δ^2, which are indicated by the vertical lines, the committor sees that region A is attracting M_2, but not M_1. Conversely, on these timescales, M_1 is stable.

B. Damped driven pendulum

In the study of dynamical systems using Markov chain approximations, we introduce discretization noise. This turns stable states into metastable states. This example shows that we can use our notions of stochastic basins to extract information on the basin structure of the underlying discrete system.

The following system of equations describes the dynamics of a damped driven pendulum [2] and is used in classical power grid models to model a single generator [1]:

$$\begin{aligned}\dot\phi &= \omega,\\ \dot\omega &= -\alpha\omega + P - K\sin\phi,\end{aligned} \tag{36}$$

The parameter values are $\alpha = 0.1, K = 1$, and $P = 0.5$. The system has a stable fixed point at $(\phi, \omega) = (\sin^{-1}\frac{P}{K}, 0)$ and a stable limit cycle at approximately $(\phi, \omega) \approx (\phi, 5)$; compare Fig. 3 for a plot of the phase space.

To compute ε-absorption stability, first we have to transform the ordinary differential equation (ODE) into a discrete dynamical system. For a fixed time step τ the flowmap $\varphi(x_0) = \varphi(\tau, x_0) = x(\tau)$ gives the value $x(\tau)$ at time τ of a solution $x(t)$ of the ODE with initial condition x_0. Then $\varphi : X \to X$ defines a discrete dynamical system and we can

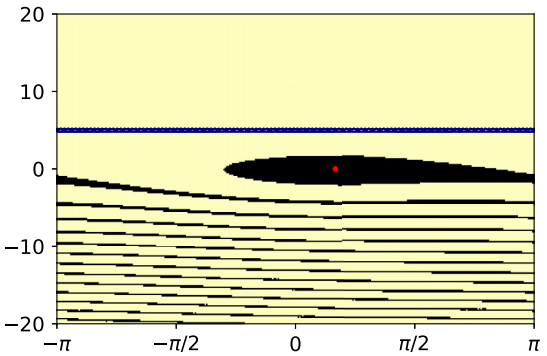

FIG. 3. Basins of attraction with corresponding fixed point (red) and limit cycle (blue) of the damped driven pendulum at parameter values $\alpha = 0.1$, $K = 1$ and $T = 0.5$.

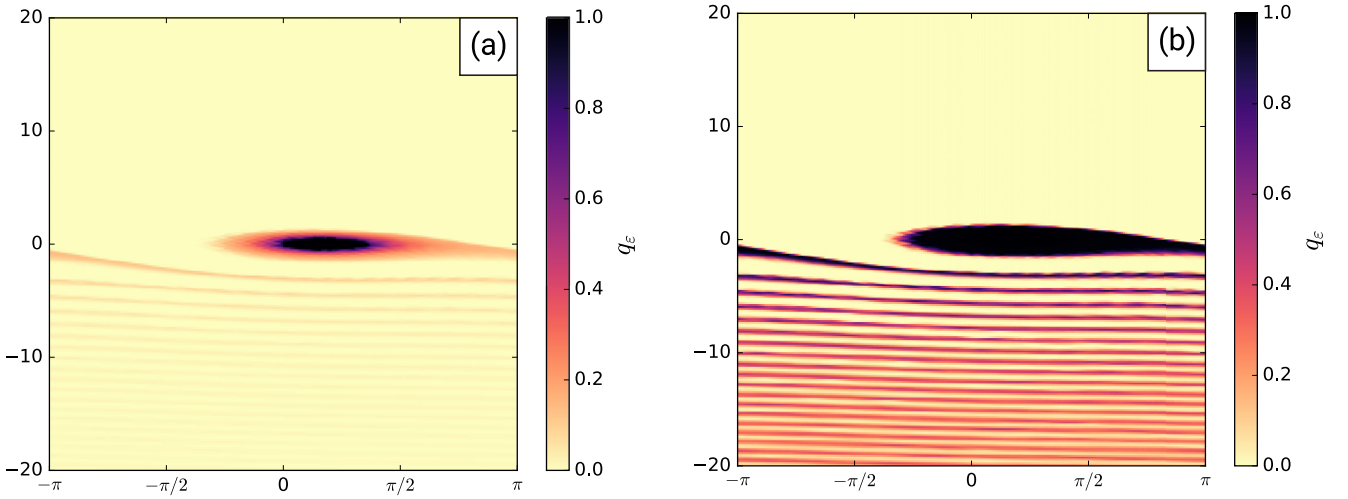

FIG. 4. ε-committors of the metastable set around the fixed point. The absorption rates are $\varepsilon = 0.1$ and $\varepsilon = 0.001$ with corresponding ε-absorption stability values of $b_\varepsilon = 0.017$ and $b_\varepsilon = 0.123$. For $\varepsilon = 0.01$ the whole basin of the fixed point is detected and b_ε is close to the basin stability value of the original system. If ε is decreased further, then the basin stays qualitatively unchanged, while the basin stability value is even better approximated.

construct the Perron-Frobenius operator and its Ulam approximation according to Appendix A 5.

For Ulam's method we use 256×256 regular square boxes on the state space $[-20, 20]\times[-\pi, \pi]$, such that the resulting transition matrix has dimension $65\,536\times65\,536$. In every box 1000 initial conditions are initiated uniformly on random and numerically integrated for the time step $\tau = 1$ to obtain the transition probabilities between boxes.

The discretization by Ulam's method introduces discretization diffusion in the system and thereby destroys the stability of the attractors, in particular of the stable fixed point, since trajectories in its basin spiral only slowly toward it, and therefore it is possible that they enter a box centered outside the fixed points' basin. However, a metastable set remains in the vicinity of the fixed point. By analyzing this set we can determine the basin of attraction of the original fixed point. If we choose less boxes for our discretization method, then the resulting discretization noise increases and metastability of the set around the fixed point decreases until its relation to the deterministic behavior is lost.

Figure 4 shows the committor functions of the metastable set around the fixed point. As expected the basin of ε-absorption converges to the basin of attraction shown in Fig. 3 when the expected time horizon is increased. The ε-absorption stability value of $b_\varepsilon \approx 0.1262$ for $\varepsilon = 10^{-8}$ is in very good accordance with the classical basin stability value

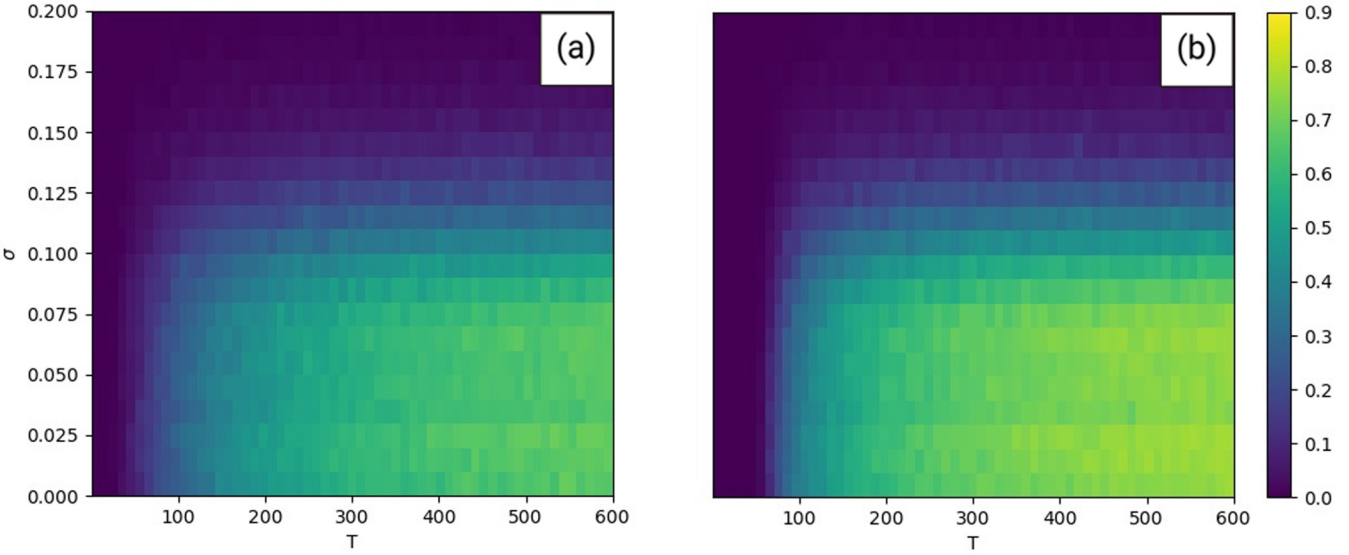

FIG. 5. Stochastic basin stability based on the mean-soujourn time (a) and the ε-committor (b) as a function of time horizon $T = \frac{2}{\varepsilon}$ and noise strength σ. Region A is the part of phase space satisfying $|\omega_i| < 0.5Hz$, close to the synchronous state. Bottom right corner converges to deterministic basin stability.

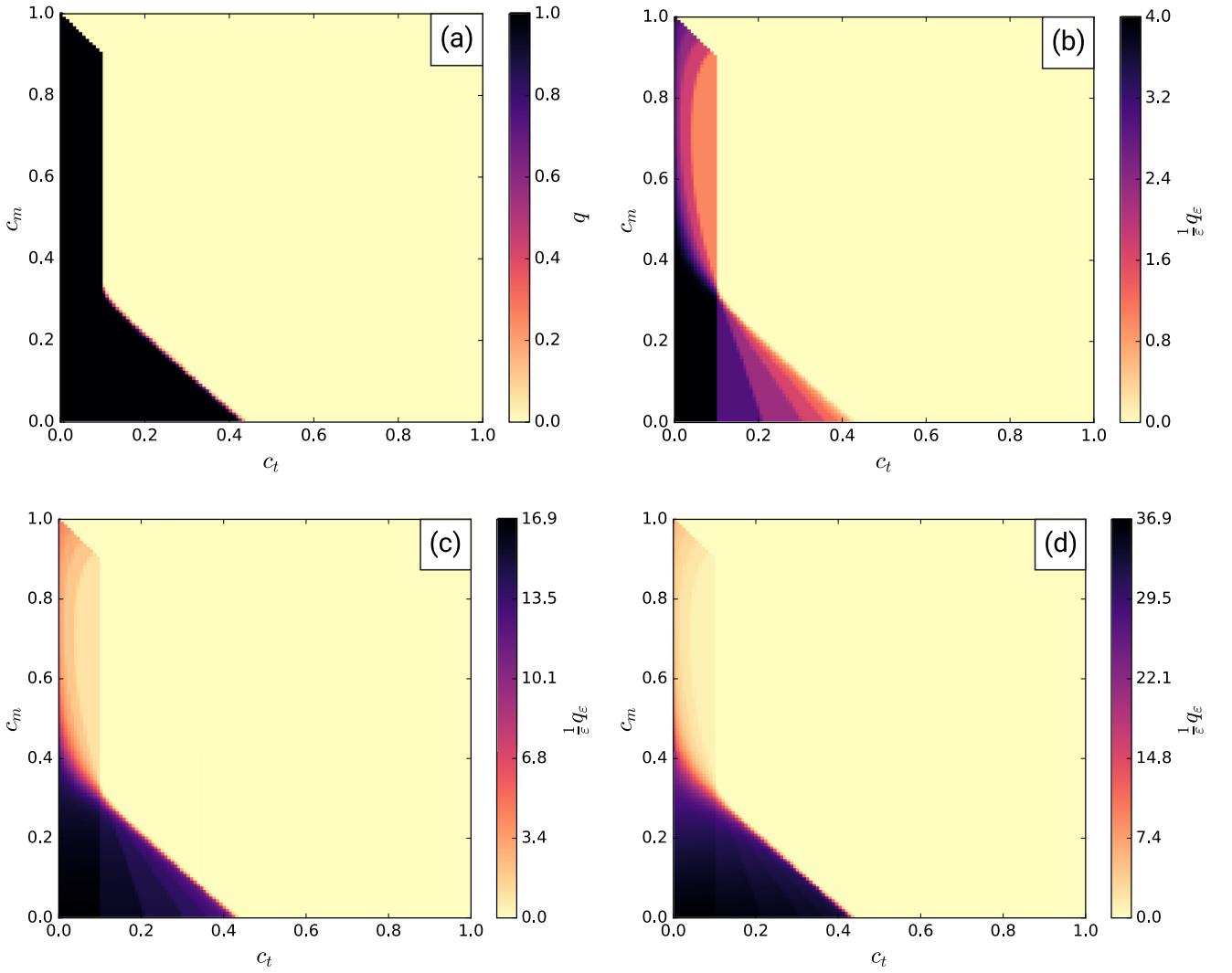

FIG. 6. (a) Classical committor function q with respect to the dead zone D and the asymptotic fixed point, cf. Eq. (11). The fraction of initial states that eventually hit the dead zone is 0.295. The other plots show the expected time the process spends in D before absorption, as described by the normalized ε-committors $\frac{1}{\varepsilon}q_\varepsilon$ for different absorption rates. In (b) $\varepsilon = 0.25$ and the ε-absorption stability $b_\varepsilon = 0.182$, (c) $\varepsilon = 0.05$ and $b_\varepsilon = 0.142$, (d) $\varepsilon = 0.0001$ and $b_\varepsilon = 0.006$. We observe that while b_ε tends to zero for ε to zero the maximum value of $\frac{1}{\varepsilon}q_\varepsilon$ converges to 36.9.

obtained by Monte Carlo integration as 0.1267 ± 0.0002. Note that the required number of function evaluations to compute basin stability up to this precision by the Monte Carlo approach is considerably higher than the number of function evalutions required to construct the transition matrix. When ε is further decreased the values of the ε-committor are expected to slowly decrease due to discretization diffusion. At a resolution of 256×256 boxes this effect is not observed since it is below numerical precision, however at a resolution of 128×128 boxes it is clearly visible and for resolutions below 64×64 boxes discretization diffusion gets too strong to draw any reliable conclusions on the systems dynamics.

Obviously the number of boxes is the main factor for determining the computational cost of Ulam's method and hence it is desirable to use as few boxes as possible. For some systems adaptive partitions may greatly reduce computational effort by using fewer partition elements [22,23].

C. A chain of oscillators

To illustrate the sampling approach for high-dimensional systems we study a chain of 16 coupled damped driven pendula, where each oscillator is subject to independent additive noise, and perturb them around the synchronous state. The equations are given by

$$\begin{aligned}
\dot{\phi}_i &= \omega_i, \\
\dot{\omega}_1 &= -\alpha\omega_1 + P_1 - K\sin(\phi_1 - \phi_2) + \sigma\dot{W}, \\
\dot{\omega}_i &= -\alpha\omega_i + P_i - K\sin(\phi_i - \phi_{i+1}) \\
&\quad - K\sin(\phi_i - \phi_{i-1}) + \sigma\dot{W}, \\
\dot{\omega}_{16} &= -\alpha\omega_{16} + P_{16} - K\sin(\phi_{16} - \phi_{15}) + \sigma\dot{W}.
\end{aligned}$$

Perturbations are from the range ±5 Hz and $\pm\pi$. We study the system with randomly drawn $P_i = \pm1$ and $K = 8$ for various levels of additive noise acting on the frequencies. This

is a surprisingly complicated model with a large number of complicated asymptotic states besides synchrony. The results for various choices of time horizon/absorption probability $T = \frac{2}{\varepsilon}$ and noise strength σ are shown in Fig. 5. The region whose basin of attraction is studied is that of all frequencies smaller than 0.5 Hz. This is qualitatively the type of constraint on the behavior of a system that one is concerned about in the context of power grid modeling.

Note that, as can be seen from the single damped driven pendulum, the region around the attractor is only metastable if noise is added to the system. Therefore, this is an example of generalized basin stability for a metastable state.

Looking at low noise, the probability to end in the region studied first increases with T. This shows the timescale on which the perturbations studied return to the metastable region. As the fixed point is the only attractor in the region, the no-noise stochastic basin stability converges to the basin stability of the attractor as T increases. With some noise added the stochastic basin stability remains close to the deterministic one, until we see the noise reach a strength where the metastability of the region studied collapses. This illustrates that our stochastic basin stabilities are a natural generalization of basin stability.

D. Anderies' model of global carbon dynamics

Anderies *et al.* [24] introduce a nonlinear conceptual model of global carbon dynamics that exhibits long transient trajectories when started in a particular region of phase space. The model equations for marine c_m, terrestrial c_t, and atmospheric c_a carbon are

$$\begin{aligned}\dot{c}_m &= \alpha_m(c_a - \beta c_m),\\ \dot{c}_t &= \mathrm{NEP}(c_a, c_t) - \alpha c_t,\\ c_a &= 1 - c_m - c_t,\end{aligned}$$

where $\alpha_m = 0.05$, $\alpha = 0.1$, and $\beta = 1$ and NEP denotes a complex, nonlinear relation between c_a and c_t, which is explained in detail in Anderies *et al.* [24]. Due to the third equation the total amount of carbon stays constant and we can consider the system on the restricted phase space $X = \{(c_m, c_t) \in [0,1]^2 \mid c_m + c_t \leqslant 1\}$. For the chosen parameters the system has a single, globally attractive fixed point and hence basin stability equals 1 by definition. Trajectories starting with low marine and terrestrial carbon stocks, i.e., $c_m + c_t \leqslant 0.4$ pass through a set where $c_t \approx 0$ before converging to the stable state.

The so-called dead zone is defined as $D := \{(c_m, c_t) \in X \mid c_t < 0.1\}$ and corresponds to a state of low terrestrial carbon stocks, i.e., when pratically all land-based vegetation and thus the basis for human life has vanished. It contains a long transient region where some trajectories spend a large amount of time before they converge to the attractor. Since the probability that the process is in D decreases monotonically for large time-horizons, we can obtain lower bounds for the expected time the process spends in D during the first ε^{-1} steps by computing the normalized ε-committor $\frac{1}{\varepsilon}q_\varepsilon$. Assuming a society is able to survive a state of low-terrestrial carbon given that vegetation recovers fast enough, the ε-committors may be used to assess which trajectories are "survivable," thus complementing the notion of "survivability" for dynamical systems recently introduced by Hellmann *et al.* [25].

To compute the ε-committors we discretize the square $[0,1]^2$ into 128×128 uniform square boxes, discard all boxes that have empty intersection with X, and compute the transition matrix according to Appendix A 5. The resulting partition has 8256 elements, where the boxes on the diagonal are triangles with half the weight of a square box.

Figure 6 shows the classical committor function of the dead zone that we introduced in Sec. II B along $\frac{1}{\varepsilon}q_\varepsilon$ for different values of ε. We chose to show $\frac{1}{\varepsilon}q_\varepsilon$ over q_ε since for transient sets the latter simply tends to zero, while the former stabilizes at the expected time the process with absorption spends in the set D, cf. Sec. IV. Given that ε^{-1} is large enough, this provides a lower bound for the expected time the original process (without absorption) spends in D, and even more, by the reasoning that we applied in Appendix C, it converges to the same value for ε to 0.

VII. CONCLUSIONS

We developed the concept of ε-committors and studied their general properties as well as asymptotic behavior. We saw that the ε-committors, as well as soujourn times, generalize basins of attraction for systems with long transients or metastable states. They can be applied to stochastic and deterministic systems likewise. The ε-committors proved especially useful in applications with an undesirable region in phase space, since they allow to compute the time the process is expected to spend in this region. Importantly we showed that the basin stability for these stochastic basins of attraction can be estimated at comparable cost to deterministic systems.

Compared to the work of Serdukova *et al.* [20], our concept of stochastic basins is entirely intrinsic and does not presuppose knowing the basin of attraction of an underlying deterministic system. This allows for a straightforward estimator for the generalized basin stability, whereas it is not known whether such an estimator exists for the definition of Serdukova *et al.* [20] (see Schultz *et al.* [26] though for an estimator for a related quantity). We define stochastic basins more generally for measurable sets in arbitrary stochastic systems given that their evolution is described by a Markov operator. The tradeoff is that our stochastic basin requires a choice of region and will in general depend on this choice. We leave working out the precise relationship between these two notions of stochastic basin to future work.

While the probabilistic formulation of the ε-committors generalize immediately to systems on continuous state spaces and for continuous-time dynamical systems, it would be interesting to also develop the appropriate PDE formulations for them, as well as for the fuzzy committors. Another interesting question is if ε-committors can be used to define metastable sets via a minimization problem. Intuitively, if a set is metastable a system that starts in a distribution in the set should be absorbed in the set again as long as ε is below the metastability timescale. This suggests that the minima of $x^T C_\varepsilon[M]x$ should contain information on the metastable sets of the system. We leave this question for future work.

VIII. SOFTWARE

The simulations were performed using Julia and Python, using the SciPy package [27]. The high-dimensional example was implemented using the DifferentialEquations.jl library [28,29] using the algorithms of Rößler [30].

ACKNOWLEDGMENTS

The authors thank Péter Koltai for many extensive helpful discussions on the use of committors in the context of basins, in which the generalized committors were defined. We also thank Jobst Heitzig and Paul Schultz for extensive comments on the final draft of this manuscript and Chris Rackauckas with help in implementing the example C. Parts of this work were funded by BMBF (CoNDyNet Grant No. 03SF0472A, CoSy-CC2 Grant No. 01LN1306A), Volkswagen Foundation (Grant No. 88462), and the Deutsche Forschungsgemeinschaft (Grant No. KU 837/39-1/RA 516/13-1).

APPENDIX A: THEORETICAL FOUNDATIONS

The following sections contain a brief introduction to dynamical systems and their operator-theoretic formulation. We will only touch upon some of the most important aspects of the theory and give references to literature. We will sometimes drop mathematical rigor for intuitive understanding and refer the interested reader to the literature and the Appendix. The purpose of this Appendix is to introduce Markov chains, and in particular mean sojourn times and committor functions, as powerful tools for studying complicated dynamical behavior.

1. Discrete time dynamical systems

Let (X, d) be a compact, metric space and let $\varphi : X \to X$ be a continuous map. The pair (X, φ) is called discrete time dynamical system. For any fixed $x \in X$ its trajectory under φ is the set $\{\varphi^k(x) \mid k \in \mathbb{N}\}$, and its *omega limit set* $\omega(x)$ is defined as

$$\omega(x) = \bigcap_{K\in\mathbb{N}} \overline{\{\varphi^k(x) \mid k \geqslant K\}}, \tag{A1}$$

where $\bar{A}$ is the closure of a set A. $\omega(x)$ is the set of all accumulation points of the trajectory of x.

$A \subseteq X$ is called *backward-invariant* with respect to φ, if $\varphi^{-1}(A) := \{x \in X \mid \varphi(x) \in A\} = A$, *forward-invariant* if $\varphi(A) = A$ and *absorbing* if $\varphi(A) \subseteq A$. It can be shown that $\omega(x)$ is closed, forward-invariant, and nonempty for every $x \in X$.

For a set $A \subseteq X$ define its *basin of attraction* B_A as

$$B_A = \{x \in X \mid \omega(x) \subseteq A\}. \tag{A2}$$

Milnor [3] calls the set B_A realm of attraction to avoid confusion with other common definitions of basin of attraction. We state without proof that for any set A its basin B_A is backward-invariant under φ, and if A is closed then B_A is Borel-measurable.

2. Stability

As we have seen above the omega limit set $\omega(x)$ is the collection of asymptotic states of a trajectory starting at x. An important question for many applications as well as for numerical modeling is whether and in what sense limit sets, or, more broadly, invariant sets, are stable. Roughly speaking a set A is locally stable if all states in a neighborhood converge to it. In other words small perturbations around A do not change the asymptotic behavior of the system. If one considers nonsmall perturbations that span the entire state space, then a global notion of stability is required.

More formally, a set $A \subseteq X$ is *Lyapunov stable* if for all neihbourhoods B' of A there exists another neihbourhood B of A from which trajectories end up in B' eventually. That is, for every $x \in B$, there is a time T after with which we have $\varphi^{T'}(x) \in B'$ for all $T' > T$.

A is *attractive* in a set $B \subseteq X$ if $\omega(x) \subseteq A$ for all $x \in B$. Further, we say that A is *locally stable* if A is Lyapunov stable and attractive in an open neighborhood U of A and that A is *globally stable* if A is Lyapunov stable and attractive in X.

From the above definition it is clear that the basin of attraction B_A of a set A is the maximal subset of X on which A is attractive. For many applications it would be desirable to know B_A exactly, to answer the question if a trajectory will converge to the same asymptotic state after an initial, possible large perturbation. Even though an exact characterization of B_A is impossible in many cases of interest, for example, due to the curse of dimensionality, it is often possible to compute the probability that a perturbed trajectory will converge back to A, as the volume of B_A under a probability measure μ modeling the perturbation. *Basin stability* of a closed set A is then defined as the volume of its basin of attraction under μ [2].

Remark. (Attractors)

Naively speaking, attractors are subsets of state space to which some initial conditions converge asymptotically. Often, an attractor is conceptualized as an invariant set that fulfills some stability property, e.g., Lyapunov stability, and is minimal in the sense that it has no proper subset with the same properties. In his definition of attractors Milnor [3] gives up the stability criterion and instead emphasizes "observability" by requiring that the corresponding basin of attraction has positive measure. According to Milnor a (minimal) attractor is a closed set $A \subseteq X$, such that

(1) B_A has positive measure, with respect to a measure μ on the Borel σ-algebra of X;

(2) there is no strictly smaller closed subset $A' \subset A$, such that $B_{A'}$ has positive measure.

In our terminology condition 1 states that B_A should have positive basin stability. Since Milnor attractors need not even be Lyapunov stable, the term "stability" might be slightly misleading. Positive basin stability implies only a positive probability that A is stable toward perturbations described by μ.

3. Transfer operators

So far we described a dynamical system by its trajectories, that is the action of a mapping φ on states $x \in X$. Equivalently, we can study the action of the composition operator or

Koopman operator $\mathrm{K}_\varphi g = g \circ \varphi$ on observables g in $L^\infty(X, \mu)$ for a given measure μ. K_φ is a bounded and linear operator.

Remark. Since X is compact, the space of continuous function from X to $\mathbb{R}$ denoted by $C := C(X, \mathbb{R})$ is contained in $L^\infty(X, \mu)$. If $g \in C$ is a fixed point of the restricted operator $\mathrm{K}_\varphi|_C$, that is $\mathrm{K}_\varphi|_C g = g$, then equivalently

$$g[\varphi(x)] = g(x) \quad \forall x \in X. \tag{A3}$$

It follows that $g[\varphi^k(x)] = g(x)$ for all $k \in \mathbb{N}$, and hence g is constant along trajectories and by continuity also on their limit sets and basins of attraction. We conclude that the fixed points of K_φ in C characterize the basin structure of φ. In particular, the constant function $\mathbb{1}_X$ is a fixed point of K_φ. For recent results on how to characterize the global stability of fixed points through the eigenfunctions of the Koopmann operator we refer the reader to Mauroy and Mezić [31].

The Koopman operator has a dual counterpart known as *Perron-Frobenius operator* which generates the evolution of probability densities f along trajectories. The Perron-Frobenius operator P_φ acts on $L^1(X, \mu)$ and is defined by requiring that for all μ-measurable sets $A \subseteq X$:

$$\int_A \mathrm{P}_\varphi f \, \mu(dx) = \int_{\varphi^{-1}(A)} f \, \mu(dx). \tag{A4}$$

Some caution has to be taken to ensure that φ and μ are compatible, for details we refer to Lasota and Mackey [32]. If f is a fixed point of P_φ, then $\mu_f(A) := \int_A f \mu(dx)$ is an invariant measure of φ, that is $\mu_f[\varphi^{-1}(A)] = \mu_f(A)$ for all measurable $A \subseteq X$.

When concerned with asymptotic behavior invariant measures generalize in many ways the notion of an attractor. This property is formulated in the famous *ergodic hypothesis* asking that for every (reasonable) observable $f : X \to \mathbb{R}$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f[\varphi^k(x)] = \int_X f dv. \tag{A5}$$

Informally speaking, the ergodic hypothesis states that the time-average of f along a trajectory should equal its space average with respect an invariant measure ν.

If we are given a measure μ that characterizes observable events, e.g., Lebesgue measure or a measure modeling a perturbation, then usually this measure is not preserved under φ. If it exists, then the invariant measure such that Eq. (A5) holds μ-almost everywhere is called *Sinai-Ruelle-Bowen (SRB) measure* [33] or sometimes *physical measure*. Unfortunately, not every system has an SRB measure and it is an active area of research to find conditions on φ that imply its existence [34]. SRB measures are useful for similar reasons as Milnor attractors—they ensure that the asymptotic states of the system are compatible with the given measure μ.

4. Markov operators and stochastic systems

The Perron-Frobenius operator introduced in the previous section is a special case of a Markov operator. From now on we understand by a *Markov operator* [32] any linear operator $M : L^1(X, \mu) \to L^1(X, \mu)$ satisfying

(1) $Mf \geqslant 0$ for all $f \in L^1, f \geqslant 0$, we say M is *positive*; and

(2) $\|Mf\|_1 = \|f\|_1$ for all $f \in L^1, f \geqslant 0$, we say M is *integral-preserving*.

Just as the Perron-Frobenius operator describes the evolution of a density in the case of deterministic dynamics, a Markov operator describes the evolution of densities under stochastic dynamics. Markov operators are closely connected to Markov processes and their transition density functions [35]. For our purposes the Markov operator framework is convenient, since it allows to characterize the statistics of the process on a density level and avoids having to deal with individual trajectories.

Another convenient property of Markov operators is that they can be approximated by Markov operators of finite-rank [36], which are just row-stochastic matrices. One such discretization scheme that goes back to an idea by Stanisław Ulam is known as Ulam's method.

5. Ulam's method for Markov operators

Let $M : L^1(X, m) \to L^1(X, m)$ be a Markov operator, where $X \subset \mathbb{R}^n$ is compact and m denotes the Lebesgue measure. Let $\mathcal{A}_h = \{A_1, \ldots, A_K\}$ be a shape-regular partition of X with mesh-size h. The basic idea of Ulam's method [35] is to obtain a coarse grained representation of the dynamics by considering only the flow of probability between partition elements. Consider the subspace $V_h \subset L^1$ spanned by the indicator functions $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_K}$. Let $Q_h : L^1 \to V_h$ be the projection onto V_h given by

$$Q_h f = \sum_{i=1}^{K} c_i \, \mathbf{1}_i \quad \text{with} \quad c_i = \int_{A_i} f dm, \tag{A6}$$

where $\mathbf{1}_i = m(A_i)^{-1} \cdot \mathbb{1}_{A_i}$ denotes the L^1-normalized indicator functions.

Lemma A.1. The discretized operator $M_h := Q_h M|_{V_h}$ is a Markov operator as well.

Proof. See Ding *et al.* [37]. ∎

Denote the matrix representation of M_h with respect to the basis $\mathbf{1}_1, \ldots, \mathbf{1}_K$ by $\hat{M}_h$. Then the matrix entries $\hat{M}_{h,ij}$ are given by the relation

$$M_h \mathbf{1}_i = \sum_{j=1}^{K} \int_{A_j} M \mathbf{1}_i \, dm \cdot \mathbf{1}_j = \sum_{j=1}^{K} \hat{M}_{h,ij} \mathbf{1}_j, \tag{A7}$$

and hence

$$\hat{M}_{h,ij} = \int_X \mathbb{1}_j M_h \mathbf{1}_i \, dm = \int_X \mathbb{1}_j M \mathbf{1}_i \, dm. \tag{A8}$$

Remark. We use the convention that matrices act by right-multiplication, i.e.,

$$M_h(f) \equiv \hat{f}^T \hat{M}_h, \tag{A9}$$

where the left-hand side signifies the operator acting on element $f \in V_h$ and the right-hand side is the matrix representation acting on the vector representation $\hat{f}$ of f.

Corollary A.2. $\hat{M}_h$ is a (row-)stochastic matrix.

Example A.1. Let M_φ be a Perron-Frobenius operator with respect to the measurable map $\varphi : X \to X$. In this case the

matrix representation of M_φ is given by

$$\hat{M}_{h,ij} = \int_X \mathbb{1}_j \, M_\varphi \, \mathbf{1}_i \, dm = \frac{1}{m(A_i)} \int_{\varphi^{-1}(A_j)} \mathbb{1}_{A_i} \, dm$$
$$= \frac{m[A_i \cap \varphi^{-1}(A_j)]}{m(A_i)}. \tag{A10}$$

We see that the entries of the matrix equal the probability that a randomly chosen state in A_i gets mapped to A_j under action of φ. Therefore, $\hat{M}_{h,ij}$ is often called *transition matrix*.

In Appendix A 3 we stated that the Perron-Frobenius operator and Koopman operator are dual to each other. Duality holds as well for the discretized operators, such that the transposed transition matrix M_h^T is an approximation of K_φ, for details see, e.g., Klus *et al.* [35].

Ulam's method is a Galerkin projection [35] and was originally developed to approximate fixed points f^* of the Perron-Frobenius operator M_φ. Thus, an important question is when the fixed points $M_{\varphi,h} f_h = f_h$ of the finite-rank approximation converge to f^* for appropriately refined partitions $\mathcal{A}_h$ of mesh size $h \to 0$. Li [38] and Ding and Zhou [39] proved convergence for certain classes of piecewise continuous maps on $\mathbb{R}^d$. Convergence of the fixed point equation in the presence of small random perturbation was shown in Froyland [40], Dellnitz and Junge [41]. Indeed, the Galerkin discretization itself may be interpreted as such a small perturbation of M that converges back to the full operator with increasing partition accuracy [40].

Since the transition matrix can be associated with a Markov chain it is sometimes referred to as Markov model [22] and was found to characterize the system's dynamical properties, even in cases where convergence of the fixed densities cannot be shown. This approach proved useful as well for approximating eigenfunctions with eigenvalues close to 1, which characterize the metastable behavior of the dynamical system [42]. Motivated by its connection to probability flows, see Eq. (A10), the transition matrix was recently interpreted as adjacency matrix of a weighted, directed graph, facilitating its analysis by tools from network theory [43,44].

6. Numerical implementation of Ulam's method

We consider the case that $M := M_\varphi$ is a Perron-Frobenius operator. From now on denote the matrix representation $\hat{M}_h$ simply as M_h. The entries $M_{h,ij}$ can be interpreted as transition probabilities from box i to box j and are usually approximated by Monte Carlo simulation. In every box A_i a large number of test points x_i^k with $k = 1, \ldots, K$ is randomly chosen, such that the transition probability can be estimated by the fraction of points that is mapped to box A_j,

$$M_{h,ij} \approx \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{A_j} \left[\varphi\left(x_i^k\right)\right]. \tag{A11}$$

It is easy to check that the resulting matrix is still stochastic and thus a numerical realization of Ulam's method [22,35,37]. For nondeterministic system, such that we are able to simulate individual trajectories, the same approach can be applied.

APPENDIX B: CONVERGENCE RESULTS

In this section we will prove convergence results for the geometric, respectively, ergodic averages of operators related to ε-committor and EMS time. We develop the theory in a general functional analytic setting since then the structure of the proofs is clearer. At the same time the results are more profound and might serve as a stepping stone for extending our concepts to transfer operators acting on infinite-dimensional spaces. Appendices B 1–B 3 establish ergodic theorems for special classes of Hilbert space operators, and Appendix B 4 focuses on the important application case of stochastic matrices. Most importantly we will see that under some assumptions the geometric, respectively ergodic averages of an operator O converge to a projection onto the fixed space of O. Recall that if O is a Koopman operator or an approximation thereof knowing its fixed space is equivalent to knowing, respectively approximating, the basin structure of the underlying dynamical system (see also Appendix A 3). These results imply that the quantities that we propose as notions of "stochastic basins of attraction," namely, ε-committors and EMS times, converge back to the classical basins of attraction in the limiting cases.

1. Ergodic theorems for contractions on a Hilbert space

This paragraph follows the approach taken by Krengel [4]. Let H be a Hilbert space, and denote the scalar product of $u, v \in H$ as $\langle u, v \rangle$. $\mathcal{B}[H]$ is the set of bounded, linear operators $O : H \to H$. Denote by O^* the dual of $O \in \mathcal{B}[H]$, such that $\langle Ou, v \rangle = \langle u, O^* v \rangle \quad \forall u, v$.

The norm of an operator $O \in \mathcal{B}[H]$ is given by

$$\|O\| = \sup_{\|v\| \leqslant 1} \|Ov\|. \tag{B1}$$

Lemma B.1. If O is a bounded, linear operator on H and O^* its dual, then

$$\|O\| = \|O^*\|. \tag{B2}$$

Example B.1. If M is a real matrix, then its dual operator is the transposed matrix M^T.

O is called *contraction*, if $\|O\| \leqslant 1$. A bounded, linear operator U is called unitary if U is surjective and preserves the scalar product, i.e.,

$$\langle Uu, Uv \rangle = \langle u, v \rangle \quad \forall u, v. \tag{B3}$$

An unitary operator is a contraction and its spectrum lies on the unit circle, see Krengel [4].

Example B.2. Any Markov operator and in particular the Perron-Frobnenius operator P_φ is a contraction on $L^1(X, \mu)$, this follows directly from the definition of a Markov operator, compare Appendix A 4. If μ is an invariant measure, then P_φ is a contraction on the Hilbert space $L^2(X, \mu)$; see Lasota and Mackey [32]. In this case the Koopman operator K_φ is a contraction on $L^2(X, \mu)$ as well [5].

The following lemmata will allow a slick proof of the classical mean ergodic theorem due to von Neumann and of a related theorem that implies the convergence of the ε-committors.

Lemma B.2. Let $O \in \mathcal{B}[H]$ be a contraction on a real or complex Hilbert space and $v \in H$. Then

$$v = Ov \quad \Leftrightarrow \quad v = O^* v. \tag{B4}$$

Proof. If for some $v \in H : \|v\|^2 = \langle v, Ov\rangle$, then $\langle v, Ov\rangle$ is real and $\langle v, Ov\rangle = \langle Ov, v\rangle$ by symmetry of the scalar product. Then we get

$$\begin{aligned}\|Ov - v\|^2 &= \langle Ov - v, Ov - v\rangle \\ &= \|Ov\|^2 + \|v\|^2 - 2\langle v, Ov\rangle \\ &\leqslant 2\|v\|^2 - 2\|v\|^2 = 0,\end{aligned} \tag{B5}$$

where we used that O is a contraction in the last line. Thus, $v = Ov$ is equivalent to $\|v\|^2 = \langle v, Ov\rangle = \langle O^* v, v\rangle$. Since O^* is a contraction as well by Lemma B.1 applying the equivalence to O^* yields the identity $v = O^* v$. ∎

We will often use the subspace $\mathrm{fix}(O) \subset H$ of O-invariant vectors

$$\mathrm{fix}(O) := \{v \in H \mid Ov = v\}, \tag{B6}$$

Obviously, $\mathrm{fix}(O)$ consists of the eigenvectors with eigenvalue 1 and is closed.

A vector u is called *orthogonal* to a subspace $V \subseteq H$, if $\langle u, v\rangle = 0 \quad \forall v \in V$. In this case we write $u \perp V$. The *orthogonal complement* $V^\perp$ of a subspace V is the set of all vectors u that are orthogonal to V.

Lemma B.3. Let $O \in \mathcal{B}[H]$ be a contraction on a Hilbert space H. Then the orthogonal complement $\mathrm{fix}(O)^\perp$ of $\mathrm{fix}(O)$ is the closure of the subspace N spanned by $\{v - Ov \mid v \in H\}$.

Proof.

$$\begin{aligned}u \perp N &\Leftrightarrow \langle u, (O - I)v\rangle = 0 \ \forall v \in H \\ &\Leftrightarrow \langle O^* u - u, v\rangle = 0 \ \forall v \in H \\ &\Leftrightarrow O^* u = u \Leftrightarrow Ou = u \Leftrightarrow u \in \mathrm{fix}(O).\end{aligned} \tag{B7}$$

Thus, N is orthogonal to $\mathrm{fix}(O)$. Since $\mathrm{fix}(O)^\perp$ is closed and contains N it contains $\overline{N}$ as well. Since a closed, linear subspace of a Hilbert space and its closure have the same orthogonal complement, we have that $\overline{N}^\perp = \mathrm{fix}(O)$ and $\mathrm{fix}(O)^\perp = (\overline{N}^\perp)^\perp = \overline{N}$. ∎

A *projection* is a linear map $P : H \to H$, such that $P^2 = P$. It induces a decomposition of $H = \ker P \oplus \mathrm{Im}\, P$ into a direct sum of its kernel and its image. If its kernel and image are orthogonal onto each other, then P is called *orthogonal projection*. The projection operator onto $\ker P$ is $Q := I - P$ and it is easy to see that $QP = PQ = 0$. Conversely, if H can be written as a direct sum of closed subspaces U and V, then every element $h \in H = U \oplus V$ can be written as $h = u + v$ with $u \in U$ and $v \in V$. The map P_U defined by $P_U h = u$, satisfies $P_U^2 = P_U$ and is called the projection onto U along V.

We are now well prepared to study the convergence of averages of powers of the operator. If O is a contraction, then we define $S_N v := S_N[O]v = \frac{1}{N}\sum_{k=0}^{N-1} O^k v$, the so-called *Cesàro averages* or *ergodic means*. Note the close connection to the expected mean sojourn times, that were introduced before.

Theorem B.4. (von Neumann mean ergodic theorem)
Let $O \in \mathcal{B}[H]$ be a contraction on a Hilbert space H. Then for every $v \in H$

$$\lim_{N\to\infty} S_N[O]v = P_{\mathrm{fix}(O)} v, \tag{B8}$$

where $P_{\mathrm{fix}(O)} : H \to \mathrm{fix}(O)$ is the orthogonal projection onto the subspace $\mathrm{fix}(O)$.

Proof. The argument is similar to the next proof, see also Krengel [4], Theorem 1.4. ∎

For a contraction O define the geometric averages $C_\varepsilon v := C_\varepsilon(O)v := \varepsilon\sum_{k=0}^{\infty}(1-\varepsilon)^k O^k v$. Since $\|O\| \leqslant 1$, it is a direct consequence of the summability of the geometric series, that for any $\varepsilon \in (0, 1]$ the operator norm of C_ε is bounded by 1 and that C_ε is linear on H. Furthermore, we have the identity

$$C_\varepsilon v = (1 - \varepsilon)C_\varepsilon(Ov) + \varepsilon v. \tag{B9}$$

Theorem B.5. (geometric mean ergodic theorem)
Let $O \in \mathcal{B}[H]$ be a contraction on a Hilbert space H. Then for every $v \in H$

$$\lim_{\varepsilon\to 0} C_\varepsilon[O]v = P_{\mathrm{fix}(O)} v, \tag{B10}$$

where $P_{\mathrm{fix}(O)} : H \to \mathrm{fix}(O)$ is the orthogonal projection onto the subspace $\mathrm{fix}(O)$.

Proof. We see immediately that $C_\varepsilon v = P_{\mathrm{fix}(O)} v = v$ for all $v \in \mathrm{fix}(O)$.

Let now $u = (O - I)v$ for some $v \in H$ then

$$\begin{aligned}C_\varepsilon u &= C_\varepsilon(Ov) - C_\varepsilon v = C_\varepsilon(Ov) - (1 - \varepsilon)C_\varepsilon(Ov) - \varepsilon v \\ &= \varepsilon[C_\varepsilon(Ov) - v].\end{aligned} \tag{B11}$$

Estimating the norm of the last term we get

$$\varepsilon\|C_\varepsilon(Ov) - v\| \leqslant \varepsilon(\|Ov\| + \|v\|) = 2\,\varepsilon\|v\|, \tag{B12}$$

and this converges to 0 for $\varepsilon \to 0$. Now let u be in the closure of $N := (O - I)H$, then there is a sequence $(u_k)_{k\in\mathbb{N}} \in N$ that converges to u, such that $u_k = (O - I)v_k$ for some $v_k \in H$. Then for every $\delta > 0$, there is $K(\delta) \in \mathbb{N}$ such that $\|u - u_{K(\delta)}\| < \delta$, and hence

$$\begin{aligned}\lim_{\varepsilon\to 0}\|C_\varepsilon u\| &\leqslant \lim_{\varepsilon\to 0}[\|C_\varepsilon(u - u_{K(\delta)})\| + \|C_\varepsilon u_{K(\delta)}\|] \\ &\leqslant \|u - u_{K(\delta)}\| + \lim_{\varepsilon\to 0} 2\,\varepsilon\|v_{K(\delta)}\| \\ &< \delta.\end{aligned} \tag{B13}$$

Since this inequality holds for all $\delta > 0$ we conclude that $\lim_{\varepsilon\to 0}\|C_\varepsilon u\| = 0$ on the closure of N, which is equal to $\mathrm{fix}(O)^\perp$ by Lemma B.3.

If $\mathrm{fix}(O)$ is a closed, linear subspace, then it is a well-known theorem that $H = \mathrm{fix}(O) \oplus \mathrm{fix}(O)^\perp$. Then we can write any $u \in H$ as $u = v + w$ with $v \in \mathrm{fix}(O)$, $w \in \mathrm{fix}(O)^\perp$ and hence $\lim_{\varepsilon\to 0} C_\varepsilon u = \lim_{\varepsilon\to 0}(C_\varepsilon v + C_\varepsilon w) = v = P_{\mathrm{fix}(O)} u$ and thus $P_{\mathrm{fix}(O)}$ is an orthogonal projection. ∎

Remark. According to the theorem the convergence of $C_\varepsilon[O]$ to $P_{\mathrm{fix}(O)}$ is pointwise. If $H = \mathbb{R}^n$, then this implies uniform convergence. For simplicity, let $\|.\|$ denote the norm induced by the standard scalar product and e_i the standard

basis vectors. Denote $D_\varepsilon := C_\varepsilon[O] - P_{\mathrm{fix}(O)}$. Then

$$\begin{aligned}\|D_\varepsilon\| = \sup_{\|x\|=1} \|D_\varepsilon x\| &\leqslant \sup_{\|x\|=1} \sum_{i=1}^{N} |x_i| \|D_\varepsilon e_i\| \\ &\leqslant N \max_{i=1,\ldots,N} \|D_\varepsilon e_i\| \\ &\leqslant N \max_{i=1,\ldots,N} 2\,\varepsilon \|f_i\| \to 0, \end{aligned} \tag{B14}$$

for $\varepsilon \to 0$, where $f_i = 0$ if $e_i \in \mathrm{fix}(O)$, or else $f_i \in H$ is such that $e_i = (O - I) f_i$. In particular the convergence is uniform if O is a contractive matrix.

Remark. We suppose that for compact, normal operators the convergence is uniform as well. A proof via the spectral theorem [45] might be possible but is, however, beyond the scope of this work.

2. Brief summary of spectral theory for Hilbert space operators

In the next section we will prove the mean ergodic theorems for another class of operators, which are not necessarily contractions. The present section introduces some of the tools needed for the proof, most notably we establish a link between the spectral radius of an operator and the convergence of its powers (Corollary B.7).

Let H be a complex Banach space and $O : H \to H$ a bounded, linear operator. The *resolvent set* $\rho(O)$ of O is the set of all $\lambda \in \mathbb{C}$, such that the operator $\lambda I - O$ is invertible with a bounded, linear inverse. Its complement $\sigma(O) := \mathbb{C} \setminus \rho(O)$ is called the *spectrum* of O. The spectrum can be split into disjoint parts, depending on the reason why the operator $\lambda I - O$ fails to be invertible.

The most important part for our purposes is the *point spectrum*,

$$\sigma_P(O) := \{\lambda \in \mathbb{C} \mid \ker(\lambda I - O) \neq \{0\}\}. \tag{B15}$$

The other parts are called the *continuous spectrum*,

$$\begin{aligned}\sigma_C(O) := \{\lambda \in \mathbb{C} \mid \ker(\lambda I - O) = \{0\},\ \mathrm{Im}(\lambda I - O) \neq H \\ \text{and } \overline{\mathrm{Im}(\lambda I - O)} = H\}, \end{aligned} \tag{B16}$$

and the *residual spectrum*,

$$\sigma_C(O) := \{\lambda \in \mathbb{C} \mid \ker(\lambda I - O) = \{0\},\ \overline{\mathrm{Im}(\lambda I - O)} \neq H\}. \tag{B17}$$

Every $\lambda \in \sigma_P(O)$ is called an eigenvalue of O and the corresponding eigenvectors are the elements of $\ker(\lambda I - O)$, which is the eigenspace of O at eigenvalue λ.

An important class of operators for which the structure of the spectrum is particularly simple and well understood are compact operators. An operator O is called *compact* if OA is relatively compact for every bounded subset $A \subset H$.

Remark. (Matrices) If H is finite-dimensional, then O is compact. In particular, every matrix is a compact operator. This is a consequence of the Heine-Borel Theorem, which states that in finite-dimensional spaces a subset is compact, if and only if it is closed and bounded.

Remark. (Compact Domain) If H is compact, then every map from H to itself is compact.

We will now state without proof a number of general results on compact operators $O \in \mathcal{B}[H]$. The proofs are omitted since they require advanced techniques that have little in common with the main subject of this paper. For details we refer to Kubrusly [45].

The so-called *Fredholm Alternative states* that the residual and continuous parts of the spectrum of a compact operator on a Hilbert space are either empty or {0}. In other words, the nonzero spectrum of a compact operator equals its point spectrum.

Theorem B.6. (Fredholm Alternative) Let $O : H \to H$ be a compact, bounded, linear operator, then

$$\sigma(O) \setminus \{0\} = \sigma_P(O) \setminus \{0\}. \tag{B18}$$

Furthermore, the spectrum is a countable set and its only possible accumulation point is 0, see Kubrusly [45], Corollary 2.20.

The *spectral radius* of an operator O is defined as

$$r(O) = \sup_{\lambda \in \sigma(O)} |\lambda|. \tag{B19}$$

The *Gelfand-Beurling formula* establishes a connection between the spectral radius of O and the norm of its powers $\|O^n\|$, it states that

$$r(O) = \lim_{n\to\infty} \|O^n\|^{1/n}. \tag{B20}$$

A proof can be found in Kubrusly [45], Theorem 2.10. This formula allows to prove that the power of an operator converges uniformly to 0 if and only if its spectral radius is strictly smaller than 1.

Corollary B.7. Let O be a bounded, linear operator on a complex Banach space, then

$$r(O) < 1 \quad \Leftrightarrow \quad \lim_{n\to\infty} \|O^n\| = 0. \tag{B21}$$

3. Ergodic theorems for a class of decomposable Hilbert space operators

We are now ready to prove the mean ergodic theorems for Hilbert space operators O that admit a decomposition into the sum of a unitary operator and an operator with spectral radius smaller than 1. An important class of such operators are stochastic matrices, as we shall see in the next section.

Theorem B.8. Let $O \in \mathcal{B}[H]$ on a Hilbert space H. Assume $H = \mathcal{U} \oplus \mathcal{V}$ is the direct sum of O-invariant closed subspaces $\mathcal{U}$ and $\mathcal{V}$, such that $U := OP_{\mathcal{U}}$ is unitary and $V := OP_{\mathcal{V}}$ has spectral radius $r(V) < 1$. Then

$$\lim_{N\to\infty} S_N[O]v = P_{\mathrm{fix}(O)}v = \lim_{\varepsilon\to 0} C_\varepsilon[O]v \quad \forall v \in H, \tag{B22}$$

where $\mathrm{fix}(O) = \ker(I - O)$ is the subspace of O-invariant vectors and $P_{\mathrm{fix}(O)}$ is an orthogonal projection.

Proof. Since the subspaces $\mathcal{U}$ and $\mathcal{V}$ are O-invariant, we have that $OP_{\mathcal{U}} = P_{\mathcal{U}}O$ and $OP_{\mathcal{V}} = P_{\mathcal{V}}O$. This implies $UV = VU = 0$ and further $O^k = (U + V)^k = U^k + V^k$. Hence, the averages $S_N[O] = S_N(U) + S_N(V)$ and $C_\varepsilon[O] = C_\varepsilon(U) + C_\varepsilon(V)$ split into two distinct terms. The mean ergodic theorems imply that

$$\lim_{N\to\infty} S_N(U)v = P_{\ker(I-U)} = \lim_{\varepsilon\to 0} C_\varepsilon(U)v. \tag{B23}$$

We show now, however, that $\lim_{N\to\infty} S_N(V)v = 0 = \lim_{\varepsilon\to 0} C_\varepsilon(V)v$.

By assumption $r(V) < 1$ and, hence Corollary B.7 implies $\lim_{k\to\infty} \|V^k\| = 0$. It follows that for all $\delta > 0$ there exists $M \in \mathbb{N}$, such that $\|V^k\| < \delta$ for all $k \geqslant M$. Then

$$\begin{aligned}\lim_{N\to\infty} \|S_N(V)\| &\leqslant \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{M-1} \|O^k\| + \lim_{N\to\infty} \frac{1}{N} \sum_{k=M}^{N-1} \|O^k\| \\ &\leqslant \lim_{N\to\infty} \left(\frac{M-1}{N} + \delta \frac{N-M-1}{N} \right) = \delta, \end{aligned} \tag{B24}$$

and since this holds for all $\delta > 0$ we have that $\lim_{N\to\infty} S_N(V) = 0$. By an analogous argument one establishes $\lim_{\varepsilon\to 0} C_\varepsilon(V) = 0$. It remains to show that $\ker(I - U) = \ker(I - O)$. By assumption for every $x \in H = \mathcal{U} \oplus \mathcal{V}$ there exist $u \in \mathcal{U}$ and $v \in \mathcal{V}$ such that $x = u + v$. Assume that $Ox = x$, then

$$\begin{aligned} Ox = x &\Leftrightarrow (U + V)(u + v) \\ &= (u + v) \Leftrightarrow Uu + Vv = u + v, \end{aligned} \tag{B25}$$

and since $Uu \in \mathcal{U}$ and $Vv \in \mathcal{V}$ this is equivalent to

$$Uu = u \quad \text{and} \quad Vv = v. \tag{B26}$$

Since V has spectral radius smaller than 1 it cannot have any fixed points except 0 and hence $v = 0$, which implies that $x = u$. Hence, $Ox = x$ if and only if $Ux = x$ or equivalently $\ker(I - U) = \ker(I - O)$. ■

Remark. The projection operator in the theorem is the orthogonal projection onto $\ker(I - U)$ along $\mathrm{Im}(I - U)$, however, in general $\ker(I - U)$ need not be orthogonal onto the subspace $\mathcal{V}$ from the theorem.

4. Ergodic theorems for stochastic matrices

Having established the mean ergodic theorems for fairly general Hilbert space operators, we now turn to most important case for the applications we have in mind, which are stochastic matrices. In this section we will introduce some basic properties of stochastic matrices and then show that they admit a decomposition like the one described in Appendix B 3. We conclude this section by proving the mean ergodic theorems for stochastic matrices.

From now on let $M \in \mathbb{R}^{n\times n}$ be a stochastic matrix. A right eigenvector of M at eigenvalue λ is a solution to the equation $Mv = \lambda v$ and similarly a left eigenvector solves $v^T M = \lambda v^T$. A left eigenvector of M is a right eigenvector of M^T and vice versa. Since M and M^T share the same characteristic polynomial, their spectra and in particular their spectral radii are the same.

Lemma B.9. M has spectral radius $r(M) = 1$.

Lemma B.10. M has at least one left eigenvector at eigenvalue 1, which is a probability distribution vector.

An eigenvalue of a matrix is called *semisimple*, if its geometric multiplicity equals its algebraic multiplicity or equivalently if the corresponding eigenspace admits an orthogonal eigenbasis (over $\mathbb{C}$). If all eigenvalues of a matrix are semisimple, then it is diagonalizable [46].

Theorem B.11. If M is a stochastic matrix, then all eigenvalues λ_i with $|\lambda_i| = 1$ are semisimple.

Theorem B.12. Let $M \in \mathbb{R}^{n\times n}$ be a stochastic matrix and let $Q \in \mathbb{R}^{n\times n}$ be invertible, such that $J = Q^{-1}MQ$ is the Jordan normal form of M, then

$$\lim_{N\to\infty} S_N[M] = \lim_{\varepsilon\to 0} C_\varepsilon[M] = P_{\mathrm{fix}(M)}, \tag{B27}$$

where $P_{\mathrm{fix}(M)}$ is a projection onto $\mathrm{fix}(M) = \{v \in \mathbb{R}^n \mid Mv = v\}$ given by

$$P_{\mathrm{fix}(M)} = Q^{-1} P_{\mathrm{fix}(J)} Q, \tag{B28}$$

and $P_{\mathrm{fix}(J)}$ is an orthogonal projection.

Proof. We assume that the reader is familiar with the Jordan normal form; a good reference is Meyer [46]. For the Jordan normal form J of M it is obvious that eigenspaces corresponding to different eigenvalues are orthogonal onto each other. As stated above the spectral radius of a stochastic matrix is 1 and all eigenvalues on the unit circle are semisimple. Hence, $J = U + V$ can be decomposed into a unitary part and a part with spectral radius smaller 1, such that $UV = VU = 0$. Since matrices are bounded, linear operators on finite-dimensional spaces, pointwise convergence implies uniform convergence. Thus, applying Theorem B.8 to J we have

$$\lim_{N\to\infty} S_N(J) = P_{\mathrm{fix}(J)}. \tag{B29}$$

Using the identity $M^2 = (Q^{-1}JQ)^2 = Q^{-1}JQQ^{-1}JQ = Q^{-1}J^2Q$ it is easy to see that

$$S_N[M] = S_N(Q^{-1}JQ) = Q^{-1}S_N(J)Q, \tag{B30}$$

and by taking limits we get

$$\lim_{N\to\infty} S_N[M] = \lim_{N\to\infty} Q^{-1}S_N(J)Q = Q^{-1}P_{\mathrm{fix}(J)}Q. \tag{B31}$$

It remains to show that $\mathrm{Im}(Q^{-1}P_{\mathrm{fix}(J)}Q) = \mathrm{fix}(M)$, which holds since

$$\begin{aligned} x^T &\in \mathrm{Im}(Q^{-1}P_{\mathrm{fix}(J)}Q) \\ &\Leftrightarrow x^T Q^{-1} P_{\mathrm{fix}(J)} Q \neq 0 \Leftrightarrow x^T Q^{-1} P_{\mathrm{fix}(J)} \neq 0 \\ &\Leftrightarrow x^T Q^{-1} J = x^T Q^{-1} \Leftrightarrow x^T Q^{-1} JQx = x^T \\ &\Leftrightarrow x^T M = x^T. \end{aligned} \tag{B32}$$

The argument for the limit $\lim_{\varepsilon\to 0} C_\varepsilon[M]$ is analogous. ■

Remark. As for any closed subspace of a Hilbert space, there exists an orthogonal projection onto $\mathrm{fix}(M)$. However, the projection $P_{\mathrm{fix}(M)}$ that we get from the theorem is, in general, not orthogonal. This happens to be so, since stochasticity of a matrix is a property that only holds with respect to a certain basis of $\mathbb{R}^n$, namely, the standard normal basis, where every basis vector corresponds to a certain state of the associated Markov chain. However, the spectrum of a linear operator is independent of the basis and the theorem is mainly a consequence of the spectral properties of M. We obtain $P_{\mathrm{fix}(M)}$ by switching to a suitable basis, such that M has Jordan normal form, which allows us to apply the results of the previous section. The resulting projection matrix $P_{\mathrm{fix}(M)}$ is sometimes called spectral projection of M at eigenvalue 1. Meyer [46] gives an explicit characterization of $P_{\mathrm{fix}(M)}$ in terms of submatrices of M.

Above we saw that the right fixed points of the transition matrix associated to a Perron-Frobenius are expected to be

almost constant on the basins, compare Appendix A 3. For general Markov chains, Deuflhard and Weber [47] give some intuition on the structure of the right 1-eigenvectors. In the ideal case of a Markov chain consisting of several uncoupled subchains, the right 1-eigenvectors will be constant on the irreducible components. If a Markov chain has several metastable states and transitions between these states are rare events, then it can be thought of as small perturbations of such an ideal chain. Deuflhard and Weber [47] show that for nearly uncoupled chains the perturbed 1-eigenvectors have eigenvalues close to 1 and are almost constant on the metastable components. They exploit this fact to approximate metastable states. However, in the presence of long transients this constant level pattern is in general not preserved and more complex algorithms are required [48,49].

APPENDIX C: DIFFERENCE ESTIMATES FOR METASTABLE STATES

This Appendix does not intend to provide thorough treatment of committor functions and EMS times in the presence of metastability but rather aims to develop our intuition of the behavior that is to be expected. Metastability in Markov chains is an extensive area of research and various related notions of metastable states and almost-invariant sets have been proposed [6,10,48]. For our purposes it will suffice to work with the straightforward idea that (left) eigenvectors with real eigenvalues close to 1 characterize metastable states. Note that the eigenvalues need to be real to avoid oscillations.

Let M be the transition matrix of a Markov chain and v an eigenvector of M at a real eigenvalue $\lambda \approx 1$, normalized such that $\|v\| = 1$. Then $\|v^T M - v^T\| = (1-\lambda)\|v\| \approx 0$ and the systems state described by v hardly changes during one iteration of the system and may be considered as invariant on short timescales.

If we consider C_ε and S_N as expected values of time averages along trajectories, then C_ε corresponds to averaging with respect to a geometric distribution with parameter ε, while S_N corresponds to averaging with respect to the equidistribution on $\{0, \ldots, N-1\}$ (cf. Sec. V). To compare both averages we align the expected values of these distribution by choosing $\varepsilon = \frac{2}{N+1}$ and write in abuse of notation $C_\varepsilon = C_{\frac{2}{N+1}} =: C_N$.

Then $s_N(v) = S_N[\lambda]v$, $q_N(m) = C_N[\lambda]v$ and for the difference term $h(\lambda, N) := \|s_N(v) - q_N(v)\|$ we have

$$h(\lambda, N) = \|S_N[\lambda]v - C_N[\lambda]v\| = |S_N[\lambda] - C_N[\lambda]| \cdot \|v\| = \left| \frac{1}{N} \frac{1-\lambda^N}{1-\lambda} - \frac{2}{N+1} \cdot \frac{1}{1-\left(1-\frac{2}{N+1}\right)\lambda} \right|. \quad \text{(C1)}$$

Elementary analytic arguments reveal that the difference vanishes for fixed λ and $N \to \infty$ as expected.

Figure 7 shows the difference terms for varied N and several eigenvalues λ, possibly far from 1. We see that for sufficiently large N the difference between EMS time and ε-committor vanishes. Further, we observe that on small timescales N such that the metastable state has not significantly decayed yet, that is $\|v^T M^N - v^T\| = 1 - \lambda^N \approx 0$, the difference term is small as well. Among our example values this effect is most significant for $\lambda = 0.999$, that is for a slowly decaying metastable states.

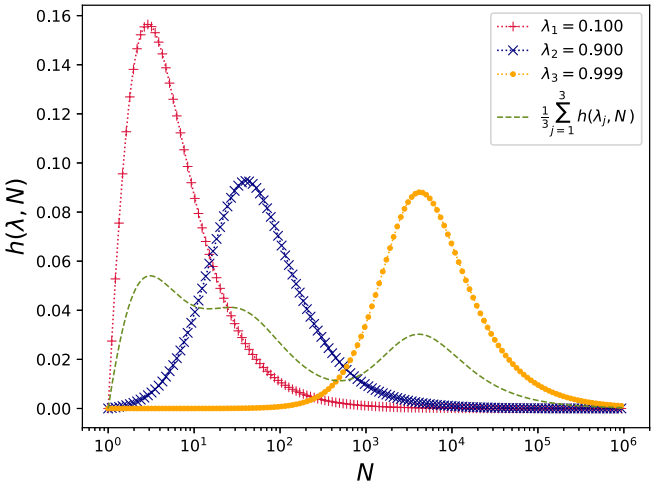

FIG. 7. Error term $h(\lambda, N)$ as described by Eq. (C1) for different values of λ and for $\varepsilon = \frac{2}{N+1}$. The green dashed line shows a difference term for a more complex metastable state given by a linear combination of three eigenvectors. In this case $\beta = \frac{1}{3}$ and $\alpha_i = 1$ for all $i = 1, 2, 3$.

If we extend our notion of a metastable state and allow $m = \beta(\sum_{i=1}^{L} \alpha_i v_i)$, where v_i is an eigenvector of M at real eigenvalue λ_i with $\|v_i\| = 1$ for all i, and $\alpha_1, \ldots, \alpha_L, \beta \in [0, 1]$, such that $\|m\| = 1$, then the difference term $\|S_N[M]m - C_N[M]m\|$ is bounded by

$$\begin{aligned} &\left\| \beta \sum_{i=1}^{L} \alpha_i v_i^T S_N[M] - \beta \sum_{i=1}^{L} \alpha_i v_i^T C_N[M] \right\| \\ &\leqslant \beta \sum_{i=1}^{L} \alpha_i \left\| v_i^T S_N[\lambda_i] - v_i^T C_N[\lambda_i] \right\| \\ &= \beta \sum_{i=1}^{L} \alpha_i h(\lambda_i, N), \end{aligned} \quad \text{(C2)}$$

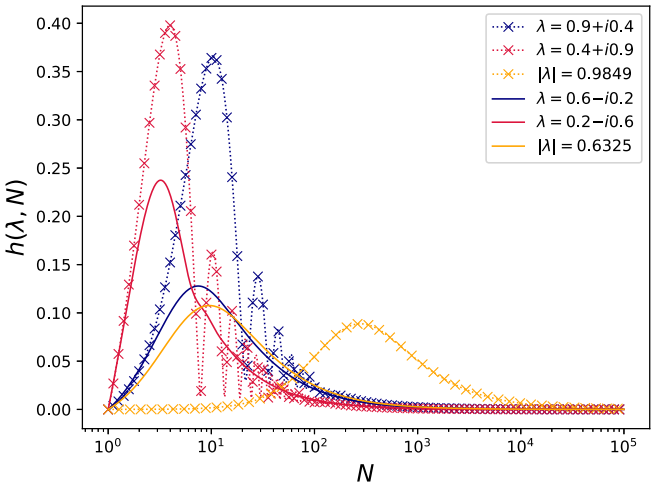

FIG. 8. $h(\lambda, N)$ for various complex λ is shown in red and blue and for comparison $h(|\lambda|, N)$ is shown in yellow. We see that for complex λ the difference term takes higher maximal values and converges faster to zero than for their absolute values.

that is by the weighted sum of the difference terms corresponding to the different eigenvectors. In Fig. 7 an example of such a combined term is plotted along the individual terms corresponding to the distinct eigenvalues. We see that the difference term for the combined metastable set decays faster than the one corresponding to the largest eigenvalue. Thus, it suffices to choose N large enough, such that $h(\lambda, N)$ is small for the largest λ to ensure that the difference term of a metastable state is small.

Remark. (Complex eigenvalues)

A generic transition matrix M is diagonalizable over $\mathbb{C}$ with possibly complex eigenvalues and eigenvectors. General linear combination of the kind $m = \beta(\sum_{i=1}^{L} \alpha_i v_i)$, with α and β as above and with complex eigenvectors v_i such that $\|v_i\|_\infty = 1$, may no longer be interpretable as metastable states due to osicillations; however, the difference terms $h(\lambda, N)$ remain valid. We observe numerically that $h(\lambda, N)$ takes larger maximal values for $\lambda \in \mathbb{C}$. However, simulations show that $h(\lambda, N)$ converges faster to zero for complex λ than for real λ with the same absolute value. Hence, to choose N large enough for the difference term of such a generalized state m to vanish we suggest to consider $h(|\lambda|, N)$ for the eigenvalue λ of largest absolute value, compare Fig. 8.

[1] P. J. Menck, J. Heitzig, J. Kurths, and H. J. Schellnhuber, How dead ends undermine power grid stability, Nat. Commun. **5**, 3969 (2014).

[2] P. J. Menck, J. Heitzig, N. Marwan, and J. Kurths, How basin stability complements the linear-stability paradigm, Nat. Phys. **9**, 89 (2013).

[3] J. Milnor, On the concept of attractor, Commun. Math. Phys. **99**, 177 (1985).

[4] U. Krengel, *Ergodic Theorems* (Walter de Gruyter, Berlin, 1985), Vol. 6.

[5] T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory* (Springer, Berlin, 2015), Vol. 272.

[6] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, Metastability and low lying spectra in reversible Markov chains, Commun. Math. Phys. **228**, 219 (2002).

[7] M. Dellnitz and O. Junge, Almost invariant sets in Chua's circuit, Int. J. Bifurcat. Chaos **7**, 2475 (1997).

[8] G. Froyland, Statistically optimal almost-invariant sets, Phys. D: Nonlin. Phenom. **200**, 205 (2005).

[9] L. Billings and I. B. Schwartz, Identifying almost invariant sets in stochastic dynamical systems, Chaos **18**, 023122 (2008).

[10] G. Froyland and K. Padberg, Almost-invariant sets and invariant manifolds-connecting probabilistic and geometric descriptions of coherent structures in flows, Phys. D: Nonlin. Phenom. **238**, 1507 (2009).

[11] R. Banisch and P. Koltai, Understanding the geometry of transport: Diffusion maps for Lagrangian trajectory data unravel coherent sets, Chaos **27**, 035804 (2017).

[12] G. Froyland, N. Santitissadeekorn, and A. Monahan, Transport in time-dependent dynamical systems: Finite-time coherent sets, Chaos **20**, 043116 (2010).

[13] G. Rubino and B. Sericola, Sojourn times in finite Markov processes, J. Appl. Prob. **26**, 744 (1989).

[14] J. R. Norris, *Markov Chains*, 2 (Cambridge University Press, Cambridge, 1998).

[15] P. Metzner, C. Schütte, and E. Vanden-Eijnden, Transition path theory for Markov jump processes, Multiscale Model. Simul. **7**, 1192 (2009).

[16] P. Koltai, *Efficient Approximation Methods for the Global Long-term Behavior of Dynamical Systems: Theory, Algorithms, and Examples* (Logos Verlag, Berlin GmbH, 2011).

[17] P. Koltai, A stochastic approach for computing the domain of attraction without trajectory simulation, in *Proceedings of the 8th AIMS Conference on Dynamical Systems, Differential Equations and Applications* (AIMS, Springfield, 2011), Vol. 2, pp. 854–863.

[18] P. Koltai and A. Volf, Optimizing the stable behavior of parameter-dependent dynamical systems—Maximal domains of attraction, minimal absorption times, J. Comput. Dyn. **1**, 339 (2014).

[19] N. Neumann, S. Goldschmidt, and J. Wallaschek, On the application of set-oriented numerical methods in the analysis of railway vehicle dynamics, Proc. Appl. Math. Mech. **4**, 578 (2004).

[20] L. Serdukova, Y. Zheng, J. Duan, and J. Kurths, Stochastic basins of attraction for metastable states, Chaos **26**, 073117 (2016).

[21] A. Csenki, *Dependability for Systems with a Partitioned State Space: Markov and semi-Markov Theory and Computational Implementation* (Springer Science & Business Media, Berlin, 2012), Vol. 90.

[22] G. Froyland, Extracting dynamical behavior via Markov models, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees (Birkhäuser, Boston, MA, 2001), p. 281.

[23] M. Dellnitz and O. Junge, An adaptive subdivision technique for the approximation of attractors and invariant measures, Comput. Visual. Sci. **1**, 63 (1998).

[24] J. M. Anderies, S. R. Carpenter, W. Steffen, and J. Rockström, The topology of non-linear global carbon dynamics: From tipping points to planetary boundaries, Environ. Res. Lett. **8**, 044048 (2013).

[25] F. Hellmann, P. Schultz, C. Grabow, J. Heitzig, and J. Kurths, Survivability of deterministic dynamical systems, Sci. Rep. **6**, 29654 (2016).

[26] P. Schultz, F. Hellmann, K. N. Webster, and J. Kurths, Bounding the first exit from the basin: Independence times and finite-time basin stability, Chaos **28**, 043102 (2018).

[27] E. Jones, E. Oliphant, P. Peterson *et al.*, SciPy: Open Source Scientific Tools for Python (2001), https://www.scipy.org/citing.html.

[28] C. Rackauckas and Q. Nie, Differentialequations.jl—A performant and feature-rich ecosystem for solving differential equations in julia, J. Open Res. Software **5**, 15 (2017).

[29] C. Rackauckas and Q. Nie, Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory, Discrete Cont. Dyn. Syst. B **22**, 2731 (2017).

[30] A. Rößler, Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations, SIAM J. Numer. Anal. **48**, 922 (2010).

[31] A. Mauroy and I. Mezić, Global stability analysis using the eigenfunctions of the Koopman operator, IEEE Trans. Autom. Control **61**, 3356 (2016).

[32] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics* (Springer Science & Business Media, Berlin, 2013), Vol. 97.

[33] L.-S. Young, What are SRB measures, and which dynamical systems have them? J. Stat. Phys. **108**, 733 (2002).

[34] E. Catsigeras and H. Enrich, SRB-like measures for C0 dynamics, Bull. Pol. Acad. Sci. Math. **59**, 151 (2011).

[35] S. Klus, P. Koltai, and C. Schütte, On the numerical approximation of the Perron-Frobenius and Koopman operator, J. Comput. Dyn. **3**, 51 (2016).

[36] J. Ding and A. Zhou, *Nonnegative Matrices, Positive Operators, and Applications* (World Scientific Publishing, Singapore, 2009).

[37] J. Ding, T. Y. Li, and A. Zhou, Finite approximations of Markov operators, J. Comput. Appl. Math. **147**, 137 (2002).

[38] T.-Y. Li, Finite approximation for the frobenius-perron operator: A solution to Ulam's conjecture, J. Approx. Theory **17**, 177 (1976).

[39] J. Ding and A. Zhou, Finite approximations of Frobenius-Perron operators: A solution of Ulam's conjecture to multidimensional transformations, Phys. D: Nonlin. Phenom. **92**, 61 (1996).

[40] G. Froyland, Computing physical invariant measures, in *Proceedings of the International Symposium on Nonlinear Theory and Its Applications* (Research Society of Nonlinear Theory and its Applications (IEICE), Tokyo, 1997), Vol. 2, pp. 1129–1132.

[41] M. Dellnitz and O. Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal. **36**, 491 (1999).

[42] G. Froyland, R. M. Stuart, and E. van Sebille, How well-connected is the surface of the global ocean? Chaos **24**, 033126 (2014).

[43] E. Ser-Giacomi, V. Rossi, C. López, and E. Hernández-García, Flow networks: A characterization of geophysical fluid transport, Chaos **25**, 036404 (2015).

[44] M. Lindner and R. V. Donner, Spatiotemporal organization of dynamics in a two-dimensional periodically driven vortex flow: A Lagrangian flow network perspective, Chaos **27**, 035806 (2017).

[45] C. S. Kubrusly, *Spectral Theory of Operators on Hilbert Spaces* (Springer Science & Business Media, Berlin, 2012).

[46] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, 2000), Vol. 2.

[47] P. Deuflhard and M. Weber, Robust perron cluster analysis in conformation dynamics, Linear Alg. Appl. **398**, 161 (2005).

[48] S. Röblitz and M. Weber, Fuzzy spectral clustering by pcca+: Application to Markov state models and data classification, Adv. Data Anal. Class. **7**, 147 (2013).

[49] M. Weber and K. Fackeldey, G-PCCA: Spectral clustering for non-reversible Markov chains, ZIB-Report (15-35) (2015).